DALARNA UNIVERSITY

Working papers in transport, tourism, information technology and microdata analysis

How does data quality in a network affect heuristic solutions?

Mengjie Han
Johan Håkansson
Pascal Rebreyend
Editor: Hasan Fleyeh

Nr: 2014:08

Working papers in transport, tourism, information technology and microdata analysis
ISSN: 1650-5581

How does data quality in a network affect heuristic solutions?

- An empirical test of a location problem with different accuracies in a road network.

Authors[1]: Mengjie Han[2], Johan Håkansson, and Pascal Rebreyend

Abstract: To have good data quality with high complexity is often seen to be important. Intuition says that the higher accuracy and complexity the data have the better the analytic solutions becomes if it is possible to handle the increasing computing time. However, for most of the practical computational problems, high complexity data means that computational times become too long or that heuristics used to solve the problem have difficulties to reach good solutions. This is even further stressed when the size of the combinatorial problem increases. Consequently, we often need a simplified data to deal with complex combinatorial problems. In this study we stress the question of how the complexity and accuracy in a network affect the quality of the heuristic solutions for different sizes of the combinatorial problem. We evaluate this question by applying the commonly used *p*-median model, which is used to find optimal locations in a network of *p* supply points that serve *n* demand points. To evaluate this, we vary both the accuracy (the number of nodes) of the network and the size of the combinatorial problem (*p*).

The investigation is conducted by the means of a case study in a region in Sweden with an asymmetrically distributed population (15,000 weighted demand points), Dalecarlia. To locate 5 to 50 supply points we use the national transport administrations official road network (NVDB). The road network consists of 1.5 million nodes. To find the optimal location we start with 500 candidate nodes in the network and increase the number of candidate nodes in steps up to 67,000 (which is aggregated from the 1.5 million nodes). To find the optimal solution we use a simulated annealing algorithm with adaptive tuning of the temperature. The results show that there is a limited

[1] Mengjie Han is a PhD-student in Micro-data analysis, Johan Håkansson is an assistant professor in Human Geography, and Pascal Rebreyend is an assistant professor in Computer Science at the School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden.

[2] Corresponding author. E-mail: mea@du.se. Phone: +46-23-778000.

improvement in the optimal solutions when the accuracy in the road network increase and the combinatorial problem (low *p*) is simple. When the combinatorial problem is complex (large *p*) the improvements of increasing the accuracy in the road network are much larger. The results also show that choice of the best accuracy of the network depends on the complexity of the combinatorial (varying *p*) problem.

Key words: complex networks, *p*-median model, simulated annealing heuristics.

1. Introduction

To have good data quality with high complexity is often seen to be important. Intuition says that the higher accuracy and complexity the data have the better the analytic solutions becomes if it is possible to handle the increasing computing time. However, for most of the practical computational problems, high complexity data means that computational times become too long or that heuristics used to solve the problem have difficulties to reach good solutions. This is even further stressed when the size of combinatorial problem increases. Consequently, we often need a simplified data to deal with complex combinatorial problems.

One such often highly complex problem is the commonly investigated *p*-median location problem, which is known as NP-hard. Therefore, heuristic methods are generally used and main issues are running time and quality of the solution produced. It's why quality of the data and size are an issue. Good data reduction or aggregation is thus often necessary to have an accurate representation of the problem. The p-median location problem uses often a road network to find optimal locations on. One aspect of the data quality is thus how accurate the representation of the road network is. Data with high accuracy is often achieved by having a complex network with huge number of nodes and edges to represent the real road network. Due to long computational times which is stressed even more when the size of the combinatorial problem increases, it is important to have knowledge of how the quality in the heuristic solutions is affected of different accuracies (number of nodes) in a network. A couple of questions can be raised: Is there a threshold beyond where there is no need to increase accuracy anymore to get good solutions and is there a trade-off between increased accuracy and increased computing time depending on the size of the combinatorial problem?

The p-median location problem is well-studied (Farahani *et al*., 2012). However, most studies are not based on real road distances but on simplified ones. Francis *et al*. (2009) made an explicit review of the p-median location problem. Among the 40 published articles, about half of them are studies based on real data. From that survey it is also obvious that almost all of the distance measures are Euclidean distance and rectilinear distance. In a recent study by Carling *et al.* (2012) the performance of the *p*-median model was evaluated when the distance measure was alternated between Euclidian, network and travel time. It was shown that for region with an asymmetrical distributed population and road network due to natural barriers the choice of distance measure has affected the optimal locations, and that the use of Euclidian distance leads to sub optimal solutions.

The work in this study follows the work of Carling *et al*. (2012). In Carling *et al*. (2012) the road network was limited to 1579 nodes and there was no analysis done of the effects on the suggested solutions by varying the number of nodes in the road network. However, differences in accuracy of the road networks could also influence the optimal location of supply points.

In a discrete location allocation problem complexity varies due to the number of demand points, number of supply points to locate and/or number of nodes in a network. However the *p*-median model is NP-hard (Kariv and Hakimi, 1979) and so aggregation has often been used to reduce the size of the problem. In our study we use a real world road network which consists of about 1.5 million nodes. To our knowledge there is no study which has used a real world network with such a high accuracy applied on a discrete p-median problem. Based on that, the aim of this paper is to analyze how the optimal location solutions vary, using the p-median model, when both the size of the combinatorial problem (number of supply points) and the accuracy of the road network are alternated. The investigation is conducted by the means of a case study in a region of Sweden, Dalecarlia. The population is distributed at 15,000 weighted demand points. The road network we elaborate is from the Swedish digital road system: NVDB (The National Road Database) and it is administrated by the Swedish Traffic administration. We use the 1.5 million nodes in the road network to calculate distances between demand and supply points. To locate supply points we use aggregations of the road network ranging from 500 candidate nodes to locate on and increasing them in steps up to 67,000.

To evaluate the effects of different accuracies in the road networks on the optimal location solutions for different sizes of the combinatorial problem we compare the results from the simulations between which we have alternated both the accuracy in the road network and the number of supply points (5-50) that are located within Dalecarlia. Since the exact optimal solution is difficult to obtain, the experiments are conducted by use of a simulated annealing algorithm.

The remaining parts of this paper are organized as follows. In section 2 we discuss some relevant literature. In section 3 we present the data used. In the fourth section we present the simulated annealing methods used. In section five we present and comment results and in section six we have a concluding discussion.

2. Literature Review

The discrete *p*-median model was first introduced by Hakimi (1964). The goal with the model is to find *p* supply points which minimize the summed distances between demands and their nearest centers. This problem can be formulated as follows.

Minimize $f = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i d_{ij} x_{ij}$, subject to $\sum_{j=1}^{n} x_{ij} = 1$ and $\sum_{j=1}^{n} x_{jj} = p$, where f is the value of objective function. n is the number demand locations. w_i is the weight of each demand location. d_{ij} is the distance from demand location i to the center j. x_{ij} is a dummy variable: taking 1 if location i is allocated to center j.

Since we model our problem as a *p*-median problem, our objective function will be to minimize the value f which is the sum of all network distances between a person and the closest supply point. (d_{ij} is one for the closest location in our case). By dividing this value by the total population, we obtain the average distance between a person and its closest supply point.

To find the optimal location for *p* supply points in relation to the demand using the *p*-median model is NP-hard, Kariv and Hakimi (1979). The complexity depends both on the number of supply points to be located, the number of demand points, as well as on how distance is measured.

Although Euclidean distance is most widely used, the network distance is in most cases more accurate in measuring the travel distance between two points (e.g. Carling *et al.* 2012). Further, a refined network should give the possibility to more accurate distance measures between two points compared to a sparser network. There are a few studies which evaluate network effects on optimal locations. Peeters and Thomas (1995) examined the *p*-median problem for different types of networks by changing the nature of the links. They found that there was a difference in optimal solutions when the links were changed but they registered no differences in computational effort.

Morris (1978) tested the linear programming algorithm for 600 random generated data sets. He generated a benchmark to simulate the effect of a road network by adding a random noise to the Euclidian distance. His conclusion was that regardless whether he was using the pure Euclidian distance or the simulated networks he was able to solve the problem, implying that the choice of distance measure is not significant. However, the data set were very small and it was only a simulated network with values close to the Euclidian distance. Further he did not really evaluate the effect of the choice of the distance measure to the quality of the solutions.

Schilling *et al*. (2000) examined the Euclidean distance, network distance and a randomly generated network distance. Their conclusion is that it is much easier for the Euclidean and network to obtain the optimal solution and with less computational effort. However, the problem is small scale and they did not provide the effect of network with different numbers of nodes in the networks. In our study we are dealing with large networks and we systematically alternate the number of nodes in it to evaluate the quality in the optimal solutions.

In a recent study Avella *et al*. (2012) tested a large size *p*-median problem using a new heuristic based on Lagrangean relaxation. The number of nodes varies from 3,038 to 89,600. They compared their computational results to the results found by Hansen *et al*. (2009) under 4 instance sets (from Birch and TSP library). The largest data set is Birch 1. The Birch data set are synthetically generated, designed to test clustering algorithms. Birch 1 and 3 differ in two significant ways. Birch 1 is the largest data set used (89,600 nodes) and it consists of symmetrical distributed demand points and nodes in the network which is also organized in tight clusters. Birch 3 consists of up to 20,000 nodes and the demand points and the nodes in the network are more asymmetrically distributed and the clusters also vary more in their characteristics. They found that the new heuristic is fast and efficient. They also showed that the quality of the optimal solutions was quite different when Birch 1 was used compared to when Birch 3 was used. Instances of type Birch 3 also took longer computing time to be solved. Larger instances exhibit worse results. However, they did not consider a real world network, when the number of nodes in the network is alternated systematically.

3. Data

3.1 Demand Points and supply points

The demand points represent the distribution of the population's residence in Dalecarlia. In this study we use the population in 2002. The figures are public produced and controlled data from Statistics Sweden (www.scb.se). The populations' residents are registered on 250 meter by 250 meter squares. We generalize each square is to its central point. Each point is then weighted by the number of people living in each square. The populations' residence location is represented by 15,729 weighted points. In total 277,725 lived in Dalecarlia during the study year. The distribution of the residents is shown in Figure 1a. The figure illustrate that the population in the region is asymmetrically distributed. The majority of residents live in the southeast corner, while the remaining residents are primarily located along the county's major rivers and lakes.

Overall, the region is not only non-symmetrical distributed, but it is also sparsely populated with an average of nine residents per square kilometer (the average for Sweden overall is 21).

Figure 1b shows some important features of the natural landscape in Dalecarlia. Firstly it is shown that the altitude in Dalecarlia vary a lot. From the south east corner with altitude below 200 the altitude increase in general towards north east. Secondly it is shown that a major river (Dalecarlia River) and some large lakes also act as natural barriers. Clearly, when comparing the distribution of the population (Figure 1a) with the natural barriers (Figure 1b) there is a correlation.

Concerning the supply points, in this study, we search for optimal locations for *p* equal 5, 10, 15, 20, 25 30, 35, 40, 45 and 50.

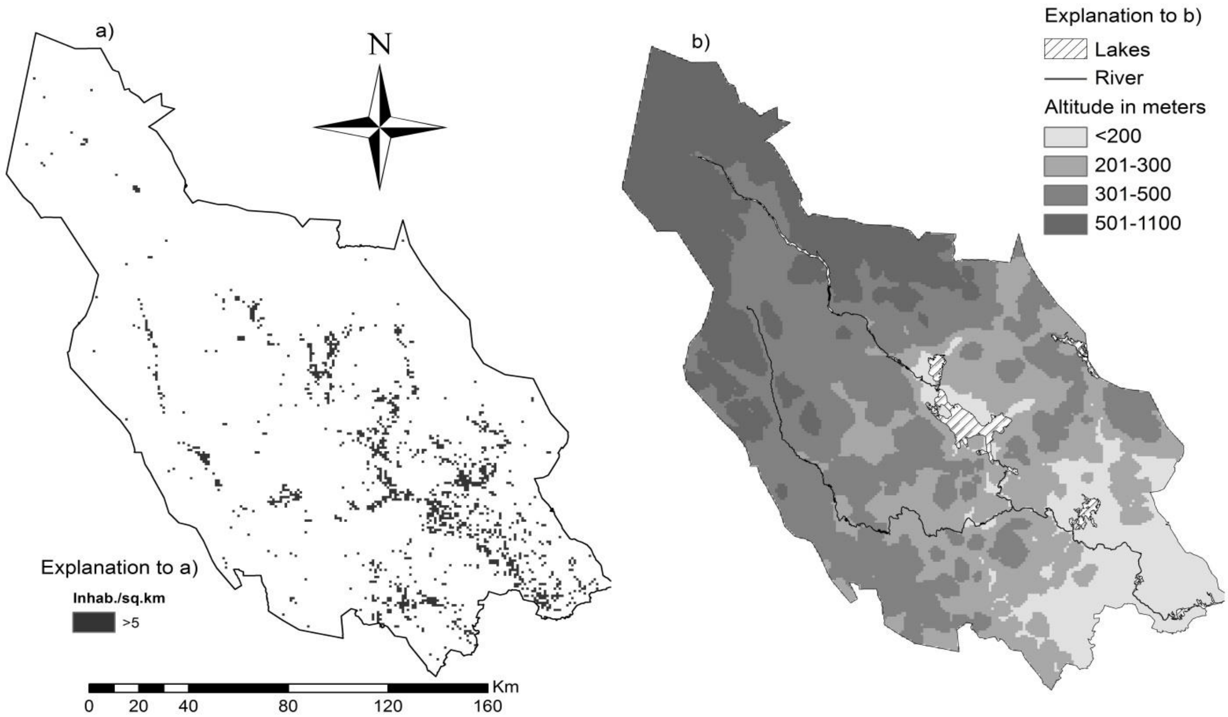

Figure 1. The distribution of the population on 1 by 1 km squares (a) and natural landscape (b) in Dalecarlia.

3.2 The Road Network

The road network used is the 2011 national road database (NVDB) for Dalecarlia. NVDB was formed in 1996 on behalf of the government. It is organized and updated by the National Transport Administration (Trafikverket) in Sweden. In total the road network for Dalecarlia contains about 1.5 million nodes and 1,964,801 segments. The total length is 39,452 kilometers. The average distance between the nodes in NVDB is about 40 meters. The minority of the nodes is nodes in intersections or at points where roads starts or begin. Most nodes describe the geographical shape of the road and by that they give a precise description of the length of the road. We use this network to calculate the distance between the demand points and the closest supply point. To do so we use the Euclidian distance to identify the closest node on the road network. Then we add the shortest network distance. To find the shortest network distance the Dijkstra algorithm has been used (Dijkstra 1959).

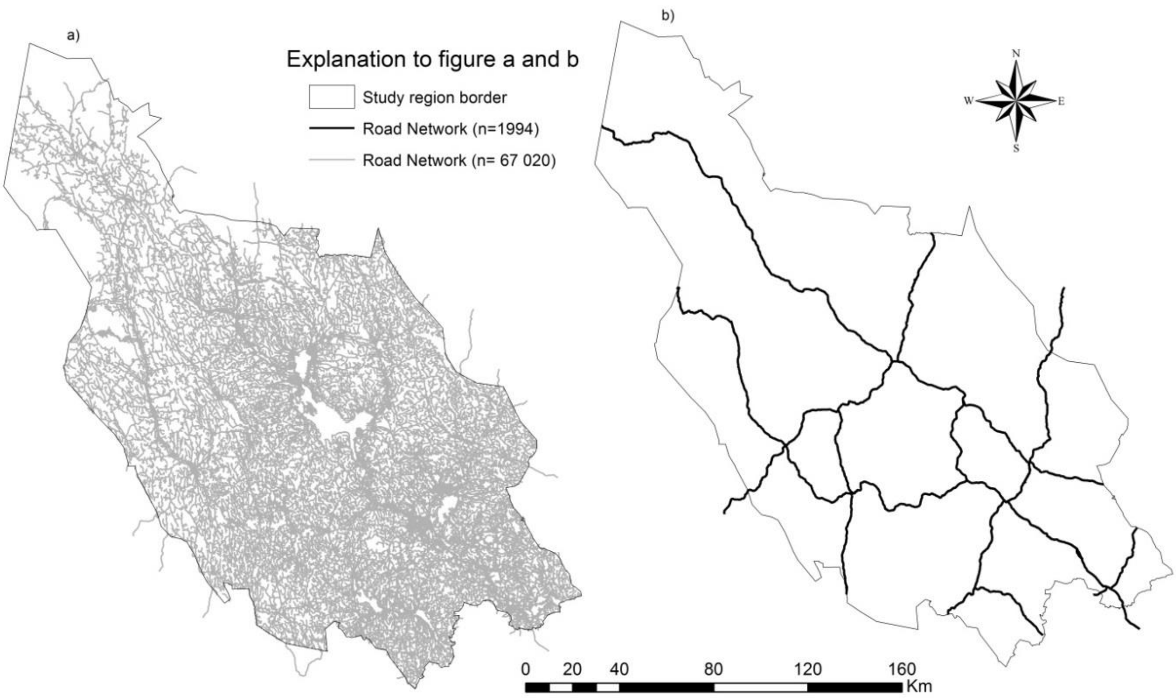

Figure 2. All roads in a dense network with 67,020 candidate nodes (a) and all roads in a sparse network with 1,994 candidate nodes (b) in Dalecarlia

To identify the candidate nodes to locate on we select one node in each 500 by 500 meter square in which the roads pass through. By reducing the number of nodes within a square an in-built location error occurs. However by selecting the center of the square as the representative node the maximum location error due to this could be 354 meters in Euclidian metric. Finally we used at the most 67,020 nodes in the road network as candidate nodes to locate on. (see Figure 2a).

NVDB is divided into 10 different categories according to the quality of the roads (see Table 1). To alternate the accuracy in the road network we used those road classes. In Dalecarlia there is just one road (class 0) which is a European highway. For this reason, class 0 roads are merged into class 1 in this study. By just taking into account the largest roads (class 0 and 1) the set of candidates to find an optimal location of a supply point are as many as about 2000 nodes distributed in a rather sparse network (see Figure 2b). This is still quite large; so to decrease the accuracy in the road network further we add two new classes which consist of 500 and 1000 candidates to locate supply points in. We select these candidates randomly from candidates in class 0 and 1. From Table 1 we can see that average distance between the candidate nodes varies rather little when the road classes 0 to 9 are concerned. However, the average distances between candidate nodes become significant longer when the accuracy in the road network is decreased further.

Figures 2a and 2b illustrate that the road network becomes denser and more homogenous in areas in the region’s southeast corner. In the southeast and in the center of the region, a sparse network of larger roads supplements the smaller roads. From Figure 2a it is obvious that the smaller local roads and streets are oriented to the larger roads. It is also evident that the smaller roads make the road network more homogenous when it comes to its distribution in the region.

Table 1. Number of candidate nodes, road length and average road distance between candidate nodes with different road classes on the road network in Dalecarlia.

Road classes	Number of nodes	Length (km)	Meters between Candidate nodes
0 to 9	67020	39454	588
0 to 8	45336	23086	509
0 to 7	20718	10964	529
0 to 6	12552	5631	449
0 to 5	12417	5479	441
0 to 4	6735	2923	434
0 to 3	3926	1725	439
0 to 2	2909	1299	446
0 to 1	1994	883	443
0 to 1 (randomized)	1000	883	883
0 to 1 (randomized)	500	883	1766

4. Simulated Annealing

4.1 Algorithm

Since the *p*-median problem is NP-hard, for large number problems, the exact optimal solution is difficult to obtain. That is why there are only a few studies examining the exact solutions (Hakimi, 1965; Marsten, 1972; Galvão, 1980; Christofides and Beasley, 1982). Our purpose in this paper is not to search for the best optimal solutions or best algorithms. What we focus on is to examine the solution patterns that can be easily obtained by the heuristic algorithms. Most studies regarding p-median problem use heuristics and meta-heuristics (e.g. Kuehn and Hamburger, 1963; Maranzana, 1964; Rahman and Smith, 1991; Rolland et al., 1996 Crainic, 2003; and Ashayeri, 2005). In our case, the cost of evaluating a solution is rather high therefore we focus on an algorithm which tries to keep the needed number of evaluated solutions low. This excludes, for example, algorithms such as the Genetic and to some extent Branch and Bound algorithms.

Another sub-class in meta-heuristics is the simulating annealing method, which we will use in this paper (e.g. Kirkpatrick 1983, Chiyoshi and Galvão, 2000; Al-khedhairi, 2008; and Murray and Church, 1996). This randomized algorithm has been chosen due to its flexibility and its ease of implementation. Even though the complexity of the problem increases, simulating annealing performs efficiently than the other algorithms. Al-khedhairi (2008) gave the general SA heuristic procedures.

SA starts with a random initial solution s, a choice of a control parameter named the initial temperature T_0, and the corresponding temperature counter $t = 0$. The next step is to improve the initial solution. The counter of the number of iterations is initially set as $n = 0$ and the procedure is repeated until $n = L$, where L is the pre-specified number of iterations of the algorithm. A neighborhood solution s' is evaluated by randomly exchanging one supply point in the current solution to the one not in the current solution. The difference, Δ, of the two values of the objective function is evaluated. We replace s by s' if $\Delta < 0$, otherwise a random variable $X \sim U(0,1)$ is generated. If $X < e^{\left(-\frac{\Delta}{T}\right)}$, s still replaces s'. The counter is updated as $n = n + 1$ whenever the replacement does not occur. Once n reaches *L*, the temperature counter is updated as $t = t + 1$ and *T* is a decreasing function of *t*. The procedure stops when the stopping condition for *t* is reached.

Given *p* we start the simulated annealing by randomly selecting points to locate the supply points. We then randomly select one of the suggested supply point location sites and define a neighborhood around it. As the neighborhood we apply a square of 25 km centered on the selected site. If we have less than 50 candidates for a supply point location we increase the neighborhood by steps of 2.5 km until this criterion is satisfied. This was necessary in just a few cases.

4.2 Adaptive Tuning and Parameters

The parameters used here have been tuned after prior testing. In our study we start with the initial temperature T_0 of 400. We multiply the temperature by 0.95 at each new iteration. To avoid having our algorithm blocked in a local minimum, we have an adaptive scheme to reheat the system. If 10 times in a row we refuse a solution, we increase the temperature multiplying the temperature by 3^{β}. A suitable value of β is 0.5. Therefore, the initial value of β is 0.5 and if no solution is accepted between two updates of the temperature we increase beta β by 0.5. β will be reset to 0.5 as soon as we accept a solution. Experiments done with 2000 and 20,000 iterations have shown that for our cases 20,000 leads to significantly better results. The number of iterations has been fixed at 20,000. Our experiments have been conducted on an Intel Core2 duo E8200 cpu working at 2.66 GHz. The operating system used is Linux and programming has been done in C and compiled with gcc. It took us about 24 hours to compute 20,000 iterations.

5. Results

Table 2 shows some results from the computer experiments when different accuracy in the Dalecarlia road network for the location of a different number of supply points is alternated. The table gives information on the mean travel distance in the road network from their residence to the closest supply point for the inhabitants in Dalecarlia. Highlighted figures in the table indicate the best solution found for a given number of supply points (*p*). When *p* is set to 15 the solutions computed continue to be better until road class 3 is added to road classes 0, 1 and 2. The best solution gives an average travel distance in the complete road network from the inhabitants' homes to the closest supply point of 8.53 kilometers.

The main result which can be drawn from Table 2 is that a more complex location problem can take advantage of a more complex network. This is shown by the fact that when the number of supply points is below 20 the best solutions are found already with the accuracy given by the road classes up to two while when the number of supply points is above 20 the best solutions are found with a higher accuracy of road network.

Table 2. The mean network distance in kilometers to the closest supply point given different *p* and densities of the road network to locate on.

	Road classes in the road network										
p	500pt	1000pt	0-1	0-2	0-3	0-4	0-5	0-6	0-7	0-8	0-9
5	22.71	20.34	19.74	**19.73**	20.20	20.17	20.25	19.99	20.07	20.33	20.52
10	11.20	11.23	11.18	**11.17**	11.20	11.22	11.26	11.38	11.44	11.74	11.80
15	8.63	8.64	8.63	**8.53**	8.58	8.58	8.66	8.63	8.75	8.92	9.21
20	7.67	7.61	7.11	7.11	7.03	7.08	7.18	**6.99**	7.24	7.54	7.82
25	7.15	7.31	7.19	6.19	6.24	**6.12**	6.16	6.15	6.30	6.61	6.94
30	6.90	6.93	6.94	5.78	5.56	**5.34**	5.58	5.55	5.68	5.88	6.05
35	6.72	6.67	6.67	5.33	5.10	**4.96**	5.15	5.13	5.16	5.29	5.43
40	6.52	6.47	6.54	5.09	4.71	4.70	4.71	4.72	**4.69**	4.89	5.29
45	6.34	6.31	6.33	4.90	4.45	**4.29**	4.40	4.42	4.49	4.69	4.84
50	6.27	6.24	6.25	4.69	4.24	4.14	**4.09**	4.12	4.24	4.40	4.46

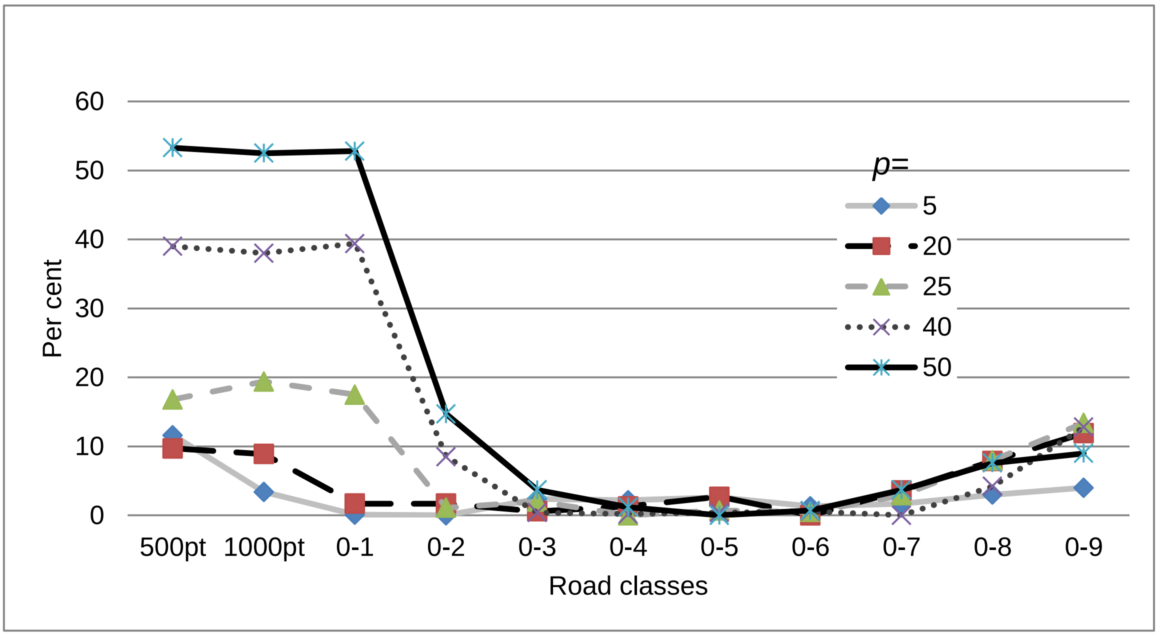

Figure 2. Variations in excess distances (in per cent) compared to the best solutions when different accuracy in the network has been used to find an optimal location on.

Figure 2 illustrates how much worse (in per cent) solutions are in relation to the best solution for different densities in the network. In the figure this is illustrated with a selection of different *p*. The conclusion is that there is more to gain in choosing the right accuracy level on the network when *p* is higher. This is clearly shown since when the number of supply point is 20 or less the worse solution found is not less than 12 per cent longer than the best one. On the other hand for location problems with more than 25 supply points the worst solution is at least 30 per cent longer than the best one.

6. Conclusions and Discussions

This paper aims to examine how the heuristic solution is affected when the accuracy of a network is varied. For a road network the accuracy varies when the number of nodes representing it varies. This study is conducted through a case study of Dalecarlia in Sweden where we alternate the accuracy of a road network, using the p-median model to find the optimal locations in it.. We use a real world road network with 1.5 million nodes of the region to calculate distances. The issue of data quality is scrutinized by systematic variation of the candidate locations from 500 to 67,000 in the network. As demand points we use the population in the region registered on squares of 250 by 250 meters. The population and the network are asymmetrical distributed in the region due to natural barriers. To scrutinize the problem we also alternate the number of *p* between 5 and 50. We use meta-heuristics to find the optimal solutions.

A major finding is that an increased quality and complexity of the road network is only necessary up to a certain level. Another finding is that when the combinatorial problem increase (*p* increases) the quality and complexity needed in the network tends to be higher.

The network used here was not constructed for the purpose of finding optimal locations within in it. The structure of it is probably suitable for a lot other of issues related to what happens on it, for instance traffic monitoring, traffic control and so on. However, in organizing this network to be suitable for the purpose of being used in location problem we turn out to have between 500 candidate nodes up to 67,000 (imply an average distance of 500 meters between the nodes) candidate nodes which are the extremes in our case. It turns out that these two extreme in the data quality of the road network were not suitable for solving the location problem here. One possible future research question could be how the road network should be arranged to be suitable for finding optimal locations within them and to design more efficient methods to reduce and aggregate data by taking into account current characteristics such as locations of nodes, population location and distribution, other available information in order to have adaptive methods.

In this study we use simulated annealing. It has obvious drawbacks. It would however be interesting to evaluate how other algorithms would perform in this kind of setting.

Acknowledgements

Financial support from the Swedish Retail and Wholesale Development Council is gratefully acknowledged. The funder has exercised no influence on this research all views expressed are solely the responsibility of the authors. The authors would also like to thank David Glémarec and Kevin Jaouen for their help in writing, testing and tuning the Simulated Annealing algorithm.

References

Al-khedhairi, A., 2008. Simulated annealing metaheuristic for solving *p*-median problem. *Int. J. Contemp. Math. Sciences* 3(28), 1357-1365.

Ashayeri, J., Heuts, R. and Tammel, B., 2005. A modified simple heuristic for the *p*-median problem, with facilities design applications. *Robotics Computer-Integrated Manufacturing* 21, 451-464.

Avella, P., Boccia, M., Salerno, S. and Vasilyev, I., 2012. An aggregation heuristic for large scale *p*-median problem. *Computers and Operations Research* 39, 1625-1632.

Carling, K., Han, M. and Håkansson, J., 2012. Does Euclidian distance work well when the *p*-median model is applied in rural areas? *Annals of Operation Research* 201(1), 83-97.

Chiyoshi, F. and Galvão, R.G., 2000. A statistical analysis of simulated annealing applied to the *p*-median problem. *Annals of Operations research* 96, 61-74.

Christofides, N. and Beasley, J., 1982. A tree search algorithm for the *p*-median problem. *European Journal of Operational Research* 10(2), 196-204.

Crainic, T., Gendreau, M., Hansen, P. and Mladenović, N., 2003. Parallel variable neighborhood search for the *p*-median. *Les Cahiers du GERAD G* 4.

Dijkstra, E. W., 1959. A note on two problems in connection with graphs, *Numerische Mathematik* 1, 269-271.

Farahani, R.Z., Asgari, N., Heidari, N., Hosseininia, M. and Goh, M., 2012. Covering problems in facility location: A review. *Computers and Industrial Engineering* 62(1), 368-407.

Francis, R., Lowe, T., Rayco, M. and Tamir, A., 2009. Aggregation error for location models: survey and analysis. *Annals of Operations Research* 167, 171-208.

Galvão, R.D., 1980. A dual-bounded algorithm for the *p*-median problem. *Operations Research* 28(5), 1112-1121.

Hakimi, S.L., 1964. Optimum locations of switching centers and the absolute centers and medians of graph. *Operations Research* 12(3), 450-459.

Hakimi, S.L., 1965. Optimum distribution of switching centers in a communications network and some related graph theoretic problems. *Operations Research* 13, 462-475.

Hansen, P., Brimberg, J., Urošević, D. and Mladenović, N., 2009. Solving large *p*-median clustering problems by primal-dual variable neighborhood search. *Data Min Knowl Disc* 19, 351-375.

Kariv, O. and Hakimi, S.L., 1979. An algorithmic approach to network location problems. part 2: The *p*-median. *SIAM J. Appl Math* 37, 539-560.

Kirkpatrick, S., Gelatt, C. and Vecchi, M., 1983. Optimization by simulated annealing. *Science* 220(4598), 671-680.

Kuehn, A. and Hamburger, M., 1963. A heuristic program for locating warehouses. *Manage Sci* 9, 643-666.

Maranzana, F., 1964. On the location of supply points to minimize transport costs. *Operations Research* 15, 261-270.

Marsten, R., 1972. An algorithmic for finding almost all the medians of a network. Technical report 23, Center for Math Studies in Economics and Management Science, Northwestern University.

Morris, J., 1978. On the extent to which certain fixed-charge depot location problems can be solved by LP. Journal of the Operational Research Society 29, 71-76.

Murray, T.A. and Church, R.L., 1996. Applying simulated annealing to location-planning models, *Journal of Heuristics* 2, 31-53.

Peeters, D. and Thomas, I., 1995. The effect of spatial structure on *p*-median results. *Transportation Science* 29, 366-373.

Rahman, S. and Smith, D., 1991. A comparison of two heuristic methods for the *p*-median problem with and without maximum distance constraints. *Int J Open Product Manage* 11, 76-84.

Rolland, E., Schilling, D., and Current, J., 1996. An efficient tabu search procedure for the *p*-median problem. *Eur J Oper Res* 96, 329-342.

Schilling, D., Rosing, K. and Revelle, C., 2000. Network distance characteristics that affect computational effort in *p*-median location problems. *European Journal of Operational Research* 127(3), 525-536.