DALARNA
UNIVERSITY

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

Network density and the p-median solution

Författare 1: Xiaoyun Zhao
Författare 2: Kenneth Carling
Författare 3: Zhiguang Dan
Författare 4: Johan Håkansson
Editor: Hasan Fleyeh

Nr: 2013:25

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

Network density and the *p*-median solution

Xiaoyun Zhao[♦], Kenneth Carling, Zhiguang Dan, and Johan Håkansson

School of Technology and Business Studies, Dalarna University, Sweden

This version: 2013-09-16

Abstract.

The *p*-median model is commonly used to find optimal locations of facilities for geographically distributed demands. So far, there are few studies that have considered the importance of the road network in the model. However, Han, Håkansson, and Rebreyend (2013) examined the solutions of the *p*-median model with densities of the road network varying from 500 to 70,000 nodes. They found as the density went beyond some 10,000 nodes, solutions have no further improvements but gradually worsen. The aim of this study is to check their findings by using an alternative heuristic being vertex substitution, as a complement to their using simulated annealing. We reject the findings in Han et al (2013). The solutions do not further improve as the nodes exceed 10,000, but neither do the solutions deteriorate.

Keywords: *P*-median Model, Vertex Substitution, Simulated Annealing, Dense Network

1 Introduction

Location problems are generally solved in one of three spaces: continuous spaces (spatial), discrete spaces, and network spaces (Hale and Moberg, 2003). Hakimi (1964) proposed the *p*-median model in the network space and showed that it is only the nodes (Hakimi, 1965) that are needed to find an optimal solution. This paper extends the investigation on the network space to check whether the *p*-median solutions improve monotonically with denser networks implying a greater number of candidate nodes for locating facilities.

Few studies have investigated the effects of a dense road network when applying the *p*-median model. Carling, Han and Håkansson (2012a) studied the effects of distance measures in an area that possesses asymmetric distributions of road network and population. Following Carling et al. (2012a), Han et al. (2013) analyzed how the optimal solutions of the *p*-median problem varied when the density of network increased. Figure 1 shows the variations of the best solutions at a given number of candidate nodes to the best solution in the network.

They used simulated annealing to do the experiments and found that best solutions did not gradually improve when the density in the road network went beyond some 10,000 nodes. However, the non-monotonic function depicted in Figure 1 may not only be due to the density of the road network, but also the performances of the algorithm.

[♦] Xiaoyun Zhao is a PhD-student in Micro-data analysis and corresponding author: xzh@du.se, phone: +4623-778509. Kenneth Carling is a professor in Statistics, Dan Zhiguang is a MSc in Business Intelligence, and Johan Håkansson is a professor in Geography.

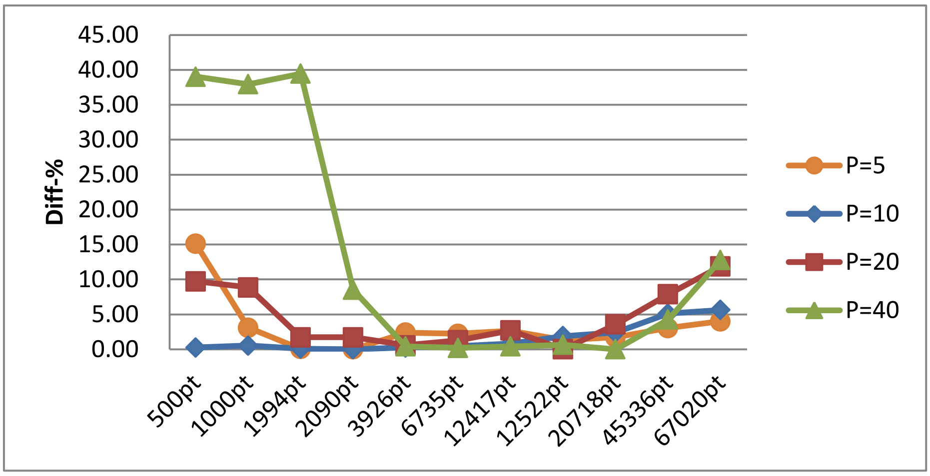

Figure 1. Variations in excess distances (in percent) compared to the best solutions for an increased number of candidate nodes. The x-axis shows the number of nodes. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to ($\frac{|\text{current solution}-\text{best solution}|}{\text{current solution}} * 100\%$). Source: Han, M. et al. (2013).

This study is mainly complementing the research done by Han et al. (2013) and further investigating whether or not the *p*-median solutions are improving monotonically with a denser network by replicating their study and adding an alternative heuristic algorithm. We also check whether or not the algorithms affect the performances of best solutions. In the study of Han et al. (2013), they stopped the checking of best solutions when the density of road network reached nearly 70,000 nodes. Based on the deteriorating results, there seemed to be no need to use a denser network considering the computational cost. As a remark, the network with 70,000 nodes is the densest one we ever encountered in our literature review.

The study is conducted with the real world network of the region Dalarna in Sweden. There are 15,729 registered demand points in the square of 250m by 250m and 1,797,939 nodes from the road network. The roads are divided into 10 road classes based on their qualities. The road class is categorized from 0 to 9 and represents different densities of the road network. Grid aggregation level of 500m by 500m is chosen to get the specific level of density we desire. The calculation of the distance between the locations of population and candidate nodes is based on the most detailed network, rather than the one after it has been aggregated. Similar to Han et al (2013), this study employs travel time as the distance measure.

2 *P*-median Model and Algorithms

2.1 *P*-median Model

The *p*-median model is one of the predominant parts in location science. Hakimi (1964) first introduced the discrete *p*-median model which became the significant mark in the development of location science; the aim of it was to find *p* facility locations which would minimize the summed distances between demand points and their nearest facilities.
Based on the formulation of Hakimi (1964), we give the *p*-median model on a network as follows:

Description variables:

$N = Totalnumberofresidentsinthespaceofinterest$

$Q = Totalnumberofdemandpointsinthespaceofinterest$

$i = Index of demand points\ (i = 1,2, \ldots \ldots, Q)$

$N_i = \ The number of residents at demand point i$

$P = Total number of potential facility locations$

$j = Index of potential facility sites\ (j = 1,2, \ldots \ldots, P)$

$w_i = The weight associated to each demand node i$

$d_{ij} = The distance between demand node i and potential facility j$

$C_{ij} = The cost of allocating i to j\ \left(C_{ij} = d_{ij} * w_i\right)$

$p = The desired number of optimal allocation$

Decision variables:

$$X_{ij} = \begin{cases} 1 & if demand at note i is allocated to facility at site j \\ 0 & otherwise \end{cases}$$

$$Y_j = \begin{cases} 1 & if a facility is located at site j \\ 0 & otherwise \end{cases}$$

The objective function is:

$$Minimize Z = \ \sum_{i=1}^{Q} \sum_{j=1}^{P} C_{ij} X_{ij} \qquad (1)$$

Subject to:

$$\sum_{j=1}^{P} X_{ij} = 1,\ \forall\ i \in (1{:}Q) \qquad (2)$$

$$\sum_{j=1}^{P} Y_j = p \qquad (3)$$

$$0 \leq X_{ij} \leq Y_j\ \ \forall\ i \in\ (1{:}Q), j \in (1{:}P) \qquad (4)$$

$$X_{ij} \in \{0,1\}\ \forall\ i \in\ (1{:}Q), j \in (1{:}P) \qquad (5)$$

$$Y_j \in \{0,1\}\ \ \forall\ j \in (1{:}P) \qquad (6)$$

The objective function (1) is to minimize the sum for all demand points of the cost to their closest facility. Constraint (2) requires that all demand points are to be assigned to only one facility point. Constraint (3) ensures that exactly *p* facility locations are to be chosen among the *P* potential ones. Constraints (4) express that no demand point *i* will be assigned to point *j* unless there is a facility. Finally constraints (5) and (6) all mean that a facility should be located at point *i* entirely or not at all.

The *p*-median problem was shown to be *NP*-hard by Kariv and Hakimi (1979); optimal solutions to large problems are difficult to obtain (Al-khedhairi, 2008). Algorithms are crucial in solving *p*-median problem. Therefore, Han et al. (2013) concluded that the best solutions of the *p*-median problem went worse when the network goes denser may be due to the performance of algorithm. More algorithms are needed to eliminate the effects of the algorithms.

Before focusing on this research question, we investigated the performances of four algorithms; greedy search, vertex substitution, Lagrangian Relaxation and simulated annealing. All of them were implemented to solve the *p*-median problem with the same data in this study for *p* equal to 7 and 11. Based on those results we chose vertex substitution as it consistently performed best of the four competitors.

2.2 Vertex Substitution (T&B)

The vertex substitution was first discussed as a local search heuristic by Teitz and Bart (1968) and we refer to it as T&B. After experimented on fifty random matrices with 20 nodes they found that this method was more stable than the partition method (Maranzana, 1964). Rosing et al. (1979) concluded that it was reasonable to use T&B to obtain a good (and potentially optimal) solution when the problem size was beyond the capabilities of optimal methods or when the requirement of true optimality was not great. Berman and Wang (2010) studied the good performance of using T&B on the *p*-median model with discrete probabilistic demand weights.

This classical interchange heuristic begins with randomly selecting an initial configuration. The current configuration will be replaced by the better solution found from its 1-neighborhood. The process iterates until the current configuration cannot be improved. So the algorithm will always terminate at an optimum, however possibly a local one. The implementation of vertex exchange can be summarized as the following steps:

1. Randomly select *p* nodes from the candidate nodes as the initial configuration *S*;
2. For the current solution *S*, calculate the objective function value OFV_S as in eq. (1);
3. Construct a set *C* of all candidate nodes not in *S*;
4. Construct the 1-neighborhood configuration of *S* (for each vertex s_i in *S* substitute s_i with every point c_i in *C*) and select a new configuration S_{new} from the neighborhood so that OFV_{new} has the smallest value of all the configurations;
5. If $OFV_{new} < OFV_S$, substitute *S* with S_{new} and goto step (3); otherwise stop the search.

T&B starts at a random state which may induce variations among the eventual results. To reduce the risk of solutions merely being local optima, we randomly select 4 initial configurations in conducting the experiment and pick the solution with the smallest value of the objective function of the 4 trials.

2.3 Simulated Annealing (SA)

Simulated annealing (SA) is one sub-class of metaheuristics that has had widespread use in the last few years. The basic idea of SA is not only accepting all the better results on the search process, but also accepting some worse results based on some specific probabilities. It is capable of solving complicated nonlinear optimization problems. It is simple to implement and it can provide high quality solutions to many problems.

Murray and Church (1996) proposed a basic SA algorithm for the *p*-median location problem. Levanova and Loresh (2004) studied the SA heuristic that used the 1-interchange neighborhood structure. Al-khedhairi, A. (2008) designed an efficient SA which found the optimal (or near optimal) solution of the *p*-median problem. Carling et al. (2012a), Han et al. (2012) implemented SA to solve the *p*-median problems in a real problem.

The performances of SA are sensitive to the values of control parameters. In this study we employ the same parameters of SA to the same real world network data as Han et al. (2013) did. The specific parameter settings and the implementation of SA are as follows:

1. Initialize the parameters: initial temperature $T = 400$; termination temperature $T_s = 0$; decrease rate of temperature $R = 0.95$; number of iterations $L = 20{,}000$; iteration counter $i = 0$; temperature increase control parameter $\beta = 0.5$ and the counter of number for no improvement $c = 0$;

2. Randomly select p points from the potential facilities set as the initial configuration S. And set the best found configuration S_{best} as the same as the initial one, i.e. set $S_{best} = S$;
3. For the current solution S, calculate the objective function value OFV_S as in formulation (1) and set the $OFV_{best} = OFV_S$;
4. Randomly select one configuration S_{new} from the 1-neighborhood of current solution S, and calculate OFV_{new};
5. Calculate the improvement on OFV_S as $\Delta OFV = OFV_{new} - OFV_{current}$;
6. If $\Delta_{OFV} < 0$, accept the new solution and set $S = S_{new}$, $OFV_S = OFV_{new}$, $i{+}{+}$, $c = 0$; otherwise accept the new solution with probability $exp(-\Delta_{OFV}/T)$, and $i{+}{+}$, $c{+}{+}$;
7. If $OFV_S < OFV_{best}$, update the optimal solution, set $S_{best} = S_{current}$, $OFV_{best} = OFV_{current}$;
8. If $c>=10$, temperature T will be increased according to $T_{new}=T_{old}*3^{\beta}$, and increase β by $\beta = \beta + 0.5$, $c = 0$;
9. If $T<=T_s$, stop the search; otherwise decrease temperature T by $T_{new} = T_{old} * 0.95$;
10. If $c > L$, stop the search; otherwise goto (4).

After testing with other tricks and parameter settings, we found that the one we use to increase the temperature works well to avoid the search being trapped in the local optimal for a long time and finally provides appealing results. Same as the T&B, the SA search also starts at a random state. We randomly select 4 different initial configurations to conduct the experiments and record the solution with the minimum objective function value.

In this paper, all the programs are coded with C and compiled using GCC; they are run under a Linux (Ubuntu) system. The testing computer has the memory of 7.9 G, CPU of Intel Core i5 3.3 GHz.

3 Data

3.1 Data Description

The data in this study is the complete digitalized representation of the road network and geo-coding of the population of Dalarna in Sweden. Figure 2 depicts the distribution of the population and the road network. Carling et al. (2012b) did a detailed description on each sub-figure; the basic data that used in this study is in accordance with their work.

Figure 2a shows the population distribution. The population data comes from Statistics Sweden (www.scb.se). The residents are registered at the points 250 meters apart in the four directions of north, west, south and east. The population number of Dalarna is 277,725 by December 2010. Regard each registry square as a point and weight each point with the number of people who live in that square. There are 15,729 demand points that represent the population in this region.

Figure 2b shows the road network with all the road classes. The data of road network of year 2011 comes from the Swedish digital road system: National Road Data Base (NVDB, www.nvdb.se). There are national roads, local roads and private streets in the Swedish road system. The total length of the road network is 39,542 kilometers.

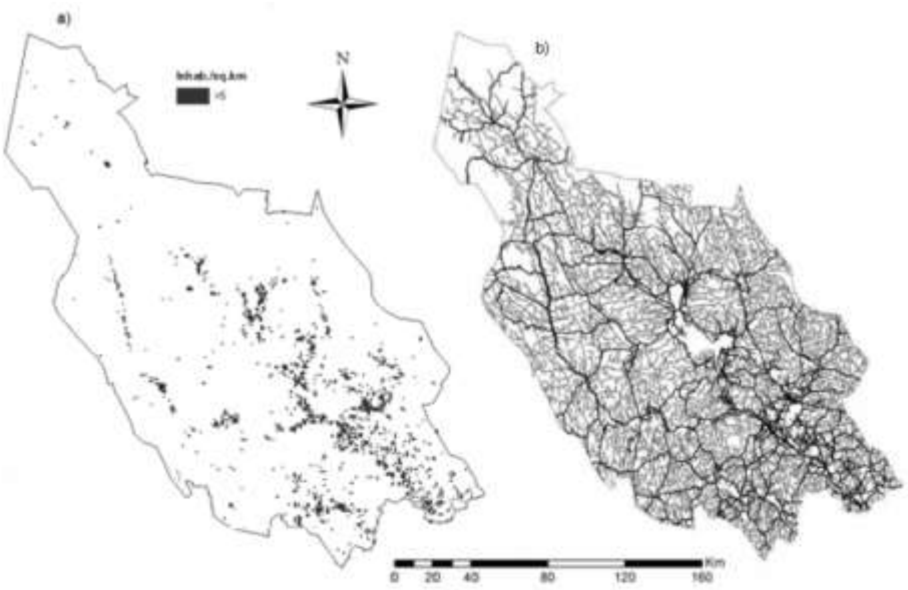

Figure 2. Map of the Dalarna region showing (a) one-by-one kilometer cells where the population exceeds 5 inhabitants, (b) national road system with local streets and subsidized private roads. Source: Carling, et al. (2012b).

3.2 Data processing

When dealing with the objective function of the *p*-median model, distance measure is one crucial aspect that should be taken into consideration. Carling et al. (2012a) first empirically investigated the consequences of different distance measures for the optimal location of multiple service centers in rural areas. They stated the shortest travel time or minimal cost along an existing network intuitively seems to be the most accurate measure for most settings, yet it is infrequently employed. One explanation they stated is that the difficulty and cost associated with collecting data on travel time. Another is the complication which arises in modeling the inherent variation in travel time. However, the data from NVDB is detailed with the information of speed limit for different roads. This paper employs travel time as the distance measure.

The residents in Dalarna region are registered at square of 250m by 250m. However, following Han et al. (2013) we represent the population by grid aggregation in 500m by 500m.

The whole road network of Dalarna is stored in two shape files, one of them includes all the information of the speeds and the directions; the other has the road classes. The road classes are also corresponding to the densities of the road network, 0 represents the sparsest network with the least nodes and 9 represents the densest network with the most nodes.

We use the c-shapefile library to process the road network file. There are 1,797,939 nodes and 1,964,801 road segments on the whole road network. The network is quite a large one for applying the *p*-median model to get relatively small *p* optimal locations. Besides, it is not necessary to set all the nodes to be the candidates for optimal locations due to the non-symmetrical sparse distribution of the population. So we use grid aggregation to select the candidates.

Table 1. Number of nodes in different road classes under the grid aggregation level of 500m by 500m on the network in Dalarna.

Road Classes	Number of Nodes
0 – 1	1548
0– 2	2237
0 – 3	3135
0– 4	5673
0 – 5	11112
0 – 6	11259
0 – 7	19556
0 – 8	44296
0 – 9	67020

The basic idea of grid aggregation is to change the density of road network by dividing it into smaller grids with the same size. In each grid at most one node is kept as a potential facility point. In this study we use three criteria to select potential facilities. First, choose the node connected with most road segments as the potential facility. Second, choose the node with the highest level as the potential facility. Third, choose the node which is the nearest one to the center of the grid as the potential facility. After the grid aggregation, the sparsest road network at road class 0 only includes the European highway and it is combined into class 1. Table 1 summarizes the number of nodes according to the highest level of the road segment that the node connects. The nodes range from 1548 to 67020.

As the coordinates of the residents do not perfectly coincide with the road nodes, we use the nearest node in the network to represent the location of the residents. This approximation partially introduces some errors in the computation. However, people usually reside at the locations where easy to get access to the transportation; some even live within the walk distance. Comparing to the time that residents spend on traveling to the nearest service center, the time of reaching to the nearest available road node can be ignored. In this study the average distance between the residents node and the nearest network node is only 62m, it does not produce much difference on the final results.

To calculate the travel time, the speed of the available road is necessary. After processing the data from NVDB, most of the roads have the speed limit of 70 km/h, which account of nearly 84% of the overall. However, there are 168 road segments with the speed of 0, for which the data are probably missing and the speed limit 70 km/h is therefore stipulated for these road segments.

The Dijkstra algorithm (Dijkstra, 1959) is used to calculate the shortest distance between each potential location node to all the population nodes. It randomly starts from a node and then calculates the distance between it and all the other nodes. After the computation, there are 9,020 nodes that do not connect with the main part of the network. Those nodes and the road segments between them are deleted. The matching between the residents' locations and the network is based on the network after deletion. The distance matrixes after calculation are stored in separate files.

4 Results and Analysis

The flowing Table 2 shows the results from the computational experiments of T&B and SA. Both the number of service centers and the density of road network are varied. We can note that the SA does not perform as well as T&B in particular in a dense network. This is mainly due to the randomness in the implementation of the SA algorithm. The results of T&B are stable after a certain density with different facility numbers. The objective function values from SA are relatively larger than T&B except when *p* is 5. As the road network becomes denser the objective function values from SA increase while those from T&B do not change substantially.

Table 2. The average travel time in seconds that demand points go to the closest service center when different facility numbers of *p* to be located on different densities road network.

Road classes	*p* =5(SA)	*p*=5(T&B)	*p*=10(SA)	*p*=10(T&B)	*p*=20(SA)	*p*=20(T&B)
0-1	964.84	964.84	586.62	586.62	419.07	418.56
0-2	957.09	961.88	582.55	581.66	387.42	384.45
0-3	957.09	961.77	582.52	581.66	386.81	382.38
0-4	957.10	957.09	582.79	581.66	382.00	376.29
0-5	957.64	956.87	582.32	580.41	383.55	371.63
0-6	956.87	956.87	582.97	580.41	385.66	371.61
0-7	959.46	956.87	585.30	580.41	383.27	371.54
0-8	960.58	956.87	591.40	580.70	394.03	371.54
0-9	961.57	956.87	593.61	580.70	400.38	371.54

In the following Figures 3-4, the road classes are employed in the x-axis. The road classes are corresponding to the nodes numbers and also the densities of the network.

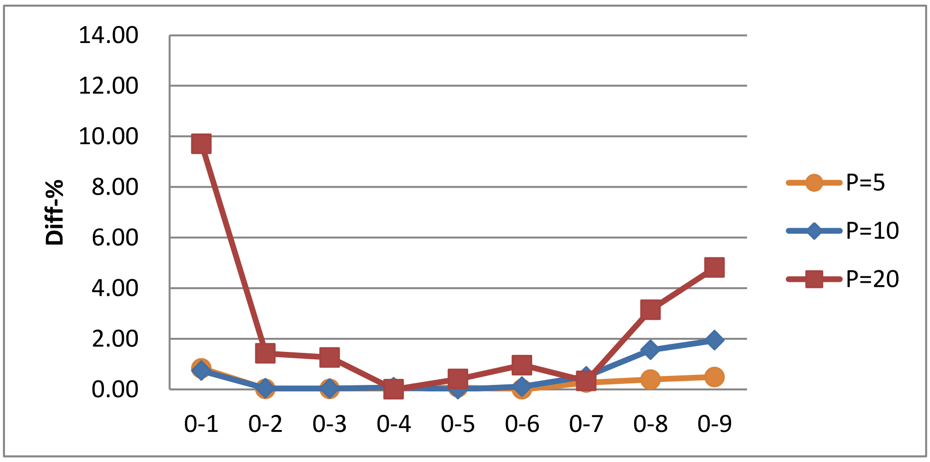

Figure 3. Variations from SA in excess distances (in percent) compared to the best solutions for an increased number of candidate nodes. The x-axis shows the number of nodes. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to ($\frac{|\text{current solution} - \text{best solution}|}{\text{current solution}} * 100\%$).

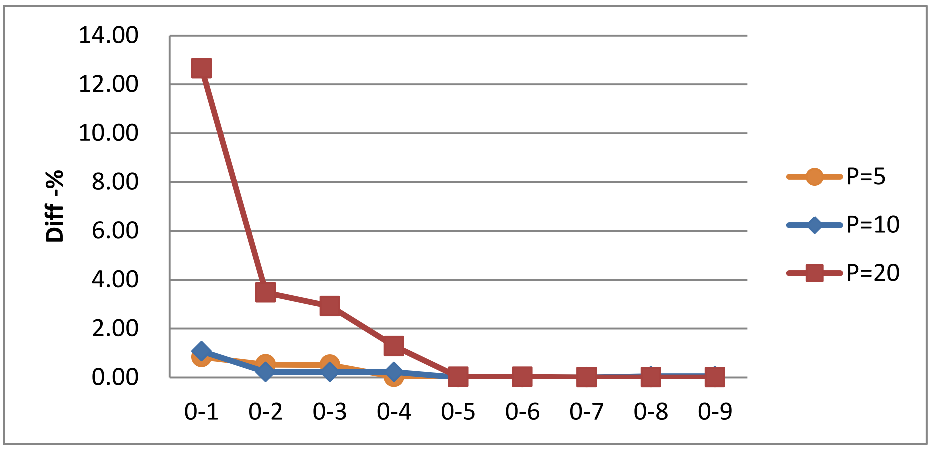

Figure 4. Variations from T&B in excess distances (in percent) compared to the best solutions for an increased number of candidate nodes. The x-axis shows the number of nodes. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to ($\frac{|\text{current solution}-\text{best solution}|}{\text{current solution}} * 100\%$).

Figure 3 shows the differences in percentage in comparison to the best solutions from SA when different densities of the road network are used. The outcome is almost identical to the results of Han et al. (2013) (cf Figure 1) and the tradeoff between how well the network is represented in a computer model and the efficiency of the SA is also apparent in Figure 3. Due to our using of travel time, instead of network distance, there is a slight difference between the results of SA here to the results of Han et al. (2013).

Figure 4 shows the differences in percentage in comparison to the best solutions when T&B is used in different densities of network. The best solutions improve monotonically and perform similarly to SA up to the road class 5. The solutions seem stable when the road classes are larger than 5, in contrast to the results of SA. Therefore, we conclude that the best results of the *p*-median model will not always be found in a denser network. It requires a proper selection of the network. However, we conclude that given an efficient algorithm and enough computing time the solution of *p*-median model improve monotonically with the density of the network.

5 Conclusion and Discussion

This paper aims at examining whether the optimal solutions of the *p*-median model improve monotonically when network becomes denser. We use a detailed network of Dalarna Province in Sweden. We aggregated the original road network into a 500m by 500m grid. The facility number alternates among 5, 10 and 20. To eliminate the algorithms' effect on the solutions of the *p*-median model for dense road network, we conduct our study by using two different algorithms being SA and T&B.

A first conclusion is that the improvement of the *p*-median solution is not monotonic along with a denser network. This is in accordance with the conclusion of Han et al. (2013). The best solution only improves before the network increase to a certain density level. Improved results can't always be derived from denser network. From Fig. 4 we note that the best solutions do not improve monotonically, nor deteriorate after the density of some 10,000 nodes. We can conclude that the choice of proper density of the network is crucial to the final results.

A second conclusion is that various algorithms perform differently in using the *p*-median model. When the network becomes denser the SA gets worse results while the T&B always gives stable results after some 10,000 nodes. The SA does not perform worse than the T&B when the network is sparser than this density.

In this study we choose 5, 10 and 20 as the facility numbers to investigate the performances of the *p*-median solutions on different densities of the road network. The road network is aggregated to the gird of 500m by 500m. More tests with larger *p* and closer aggregated approximations are needed to generalize our conclusions.

There are many other algorithms that have been used to get optimal locations from the *p*-median model. In this study the vertex substation is not able to give the solution within an acceptable time for *p* varies to 30, 40, 50 and even larger. So this would need some other algorithm to complete the study with larger p. The real world data we have used in this study is quite detailed. However, it is still a small geographical rural area compared to the whole region of Sweden. More case studies with different kinds of service centers are needed to modify the investigation at a national level.

References

1. Al-khedhairi, A. (2008). Simulated annealing metaheuristic for solving p-median problem. Int.J.Contemp. Math. Science, 3(28), 1357-1365.
2. Berman, O. and Wang, J. (2010). The network p-median problem with discrete probabilistic demand weights. Computers and operations Research, 37(8), 1455-1463.
3. Carling, K., Han, M. and Håkansson, J. (2012a). Does euclidean distance work well when the p-median model is applied in rural areas? Annals of Operations Research, 201(1), 83–97.
4. Carling, K., Han, M. and Håkansson, J. and Rebreyend, P. (2012b). Distance measure and the p-median problem in rural areas. (Working papers in transport, tourism, information technology and microdata analysis ISSN: 1650-5581, 2012:07). Dalarna University.
5. Dijkstra, E.W. (1959). A note on two problems in connection with graphs, NumerrischeMathematik 1, 269-271.
6. Hakimi, S. L. (1964). Optimal location of switching centers and the absolute centers and medians of graph. Operational Research, 12(3), 450-459.
7. Hakimi, S. L. (1965). Optimal distribution of switching centers in a communications network and some related graph theoretic problems. Operations Research, 13, 462-475.
8. Hale, Trevor S. and Moberg, Christopher R. (2003). Location Science Research: A Review. Annals of Operations Research, 123(1-4),21-35.
9. Han, M., Håkansson, J. and Rebreyend, P. (2013). How do different densities in a network affect the optimal location of service centers? (Working paper for Transportation research, Series B: Methodological ISSN: 1650-5581, 2013:15). Dalarna University.
10. Han, M., Håkansson, J. and Rebreyend, P. (2012). How does the use of different road networks effect the optimal location of facilities in rural areas? (Working papers in transport, tourism, information technology and microdata analysis, ISSN: 1650-5581, 2012:02). Dalarna University.
11. 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.
12. Maranzana, F. E. (1964). On the location of supply points to minimize transport costs. Operations Research Society, Vol. 15, No. 3, 261-270.
13. Murray, Alan T. and Church, Richard L. (1996). Applying simulated annealing to location-planning models. Journal of Heuristics,Vol.2(1), 31-53
14. Rosing, K. E., Flillsman, E.L. and Rosing-Vogelaar, H. (1979). A note comparing optimal and heuristic solutions to the p-median problem. Geogr. Analysis 11, 86-89.
15. Teitz, M. B. and Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. Operations Research, 16(5), 955-961.