

# How Does the Complexity of A Road Network Affect Optimal Facility Locations?

Xiaoyun Zhao

Microdata Analysis department, Dalarna University, SE-79 188, Falun, Sweden

Email: xzh@du.se

Pascal Rebreyend and Johan Håkansson

Computer Science department and Information Science department, Dalarna University, SE-79 188, Falun, Sweden

Email: {prb, jhk}@du.se

**Abstract**—Road network serves as a critical link in sustainable energy and environment development. To fulfill the transportation needs and to adapt to the region development, the road network is often organized hierarchically and asymmetrically with various road levels and spatial structures. The complexity of the road network therefore varies. Location models locate usually facilities on a given road network, the most complex one, and the influence from the complexity of road network in finding optimal locations is not well-studied. This paper aims to investigate how the complexity of a road network affects the optimal facility locations by applying the widely-applied  $p$ -median model. In a specific case study, the main result indicates that an increase in complexity, up to a certain level, can obviously improve the solution; the complexity beyond that level does not always lead to better solutions. A detailed sensitivity analysis of algorithm and facility number further provides insight into computation complexity and location problems from intra-urban to inter-urban.

**Index Terms**—information system, transportation system, spatial optimization, location models, heuristics, road network

## I. INTRODUCTION

Movements of people and goods take place between origins and destinations on the road network; therefore, the road network is a necessary component in facilitating these spatial movements. A facility usually serves as a destination of the movements, thus, a facility optimally located on the road network is a desired goal in many decision-making processes since it minimizes travel distance [1], [2]. In the process of urbanization, the road network expansion exerts extensive and profound influences on regional systems [3]. A complexly expanded road network is often organized hierarchically and asymmetrically with various road levels and spatial structures in order to fulfill transportation needs and to adapt to facility development. The complexity of a road network increases along with the increase of road levels and spatial structures, thus finding the optimal facility locations could become computationally complicated and time-consuming.

In locating facilities, a road network is usually given [4]. Peeters and Thomas [5] claimed that the more complex a road network, the higher probability of identifying a more optimal solution. Following which, many studies always work with a most complex road network and focus on aggregating the demand points and using a simple distance measure, such as the Euclidean distance, to simplify the problem and make it computationally treatable [6], [7]. In addition, the given road network in these studies is simulated, and the number of facility locations is limited to represent various scenarios. In using a real road network for finding optimal locations in rural areas, Carling *et al.* [1], [8] demonstrate that the level of aggregation needs to be more sophisticated and the distance measure needs to be as accurate as possible in order to avoid suboptimal locations in areas where demands and road network are asymmetrically distributed. Few studies use a real road network with changing complexity, along with the road levels and the spatial structures, in finding optimal facility locations in both intra-urban and inter-urban scenarios.

In a pre-test conducted by Rebreyend *et al.* [9], the complexity of the road network is found influential in finding the optimal location in rural areas. However, the trade-off among more optimal solutions, computational complexity and spatial distributions of facilities in intra-urban and inter-urban scenarios has not yet been studied.

This paper aims to investigate how the changes in road network complexity influence the optimal facility locations by applying the widely used  $p$ -median model. In addition, to provide further insight into computation complexity and location problems from intra-urban to inter-urban, a detailed sensitivity analysis of four algorithms and various facility numbers is conducted.

The four algorithms are: greedy search [10]; CPLEX; simulated annealing [11]; and imp-Genetic algorithm [12]. Each algorithm works in a different way in solving the  $p$ -median model. The facility numbers vary from 5 to 50 (with a common difference of 5) to simulate intra-urban and inter-urban location distributions in the case of Dalarna province, Sweden. The size of Dalarna, as well as its structure, is similar to the regions of Vermont and New Hampshire in the US.

The entire road network database provided by the official National Road Database (NVDB)<sup>1</sup> in Sweden consists of 2.6 million roads and 30 million nodes, where 0.15 million roads and 1.8 million nodes are distributed in the Dalarna region. The road network in the Dalarna region has a hierarchical structure composed of ten levels of road. The European highways are at the biggest level, representing the simplest road networks, where there is a limited number of candidate nodes for locating facilities. The local and private streets are at the smallest level and represent the most complex road network, with a huge number of candidate nodes for locating facilities.

The remaining part of the paper is organized as follows. In Section 2, we present the  $p$ -median model and some related previous studies that guide us in the model choice. Section 3 describes the data applied in the case study and settings for the sensitivity check. Section 4 illustrates the main results, and Section 5 concludes the paper.

## II. THE $p$ -MEDIAN MODEL

The  $p$ -median model was first introduced by Hakimi [13], and it has since then been widely used in transportation and location-allocation studies [14].

$$\text{Minimize } f = \sum_{i=1}^Q \sum_{j=1}^N w_i * d_{ij} * X_{ij} \quad (1)$$

Subject to:

$$\sum_{j=1}^N X_{ij} = 1, \forall i \in (1, 2, \dots, Q) \quad (2)$$

$$\sum_{j=1}^N Y_j = p, p \in N \quad (3)$$

Decision variables:

$$X_{ij} = \begin{cases} 1, \text{if demand at node } i \text{ is} \\ \text{located to facility } j \\ 0, \text{if not} \end{cases} \quad (4)$$

$$Y_j = \begin{cases} 1, \text{if a facility is located at} \\ \text{candidate site } j \\ 0, \text{if not} \end{cases} \quad (5)$$

$$\begin{cases} 0 \leq X_{ij} \leq Y_j \\ X_{ij} \in \{0, 1\} \\ Y_j \in \{0, 1\} \end{cases}, \forall \begin{cases} i \in (1, 2, \dots, Q) \\ j \in (1, 2, \dots, N) \end{cases} \quad (6)$$

$Q$  = Total number of demand points in the space of interest

$N$  = Total number of candidates for locating facilities

$p$  = Total number of potential facilities

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

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

Given that demand nodes are fixed in the network, the model finds the optimal facility locations that minimize the total travel distance for all demand points to the closest facility. Specifically, the objective function is to

minimize the sum of weighted distances between demand points and their respective nearest facilities.

The constraint (2) requires that each demand point is assigned to exactly one facility. The constraint (3) ensures that exactly  $p$  facility locations are to be chosen among the  $N$  candidates. The constraint (6) links the location variables and the allocation variables. What's more, value constraints for  $X_{ij}$  and  $Y_j$  insure that the location variables( $X$ ) and allocation variables( $Y$ ) are binary.

To find the optimal location for  $p$  facilities to fulfill the demand using the  $p$ -median model is NP-hard [15], and optimal solutions to large problems are difficult to obtain [16]. Francis *et al.* [7] conducted a literature review of the  $p$ -median model, in which about half of the 40 reviewed articles are studies based on real data. The largest number of candidate nodes in the reviewed studies was some 70, 000 candidate nodes. However,  $p$  was relatively small (<20) in those studies. The review also showed that almost all distance measures are Euclidean distance and rectilinear distance.

Many recent studies have been developed to illustrate the greater efficiency of network distance versus Euclidean distance, in the analysis of network-constrained objects or phenomena [17]. Schilling *et al.* [6] examined the Euclidean distance, network distance and a randomly generated network distance. Both Euclidean distance and network distance were found to have high computational efficiency and yield better solution quality. The problem scale in their study was, however, small and they did not study a network with different levels of candidate nodes. Although Euclidean distance is most widely used, the network distance in most cases is more accurate in measuring the travel distance between two points, since the Euclidean distance leads to suboptimal solutions in an unpredictable way. Xie and Yan [18], and Shiode [19], found that measuring distance by connecting the straight line between locations could possibly overestimate the clustering tendency of the network. The network distance is more appropriate for spatial phenomena or activities constrained by transportation networks, especially in the field of microscopic analysis [20]. Peeters and Thomas [5] examined the performance of the  $p$ -median model in different network topologies by changing the nature of the links. They found that there is a difference in optimal solutions when the links are changed, but they did not check the computational efforts in finding the optimal solutions. Following them, the topologies of the network have been studied, taking the computational effort into account ([4], [21]-[23]). However, on the one hand, these studies are mainly based on fully connected Euclidean and rectilinear networks or networks with varying numbers of radial and rectilinear arcs. On the other hand, the number of candidate nodes and links are limited, which means that the simulation cannot represent the real road network well.

The complexity of a real road network varies along with the road levels and the spatial structures, which cannot be simply represented by nodes and links in topology. Apart from Rebreyend *et al.* [9], very few

<sup>1</sup> www.nvdb.se

studies have examined the impact of varying the complexities of road network on the optimal facility locations and their spatial distributions.

### III. HELPFUL HINTS DATA AND SETTINGS FOR SENSITIVITY CHECK

#### A. Data

Dalarna is a province located in the middle of Sweden with an area of 28,189 km<sup>2</sup>. The population of Dalarna amounts to 231,934, in December 2015 (Statistic Sweden<sup>2</sup>). The population is geo-coded and registered on 250 meters by 250 meters squares. The center of each registry square represents one demand point. Each demand point is assigned a weight, corresponding to the population number in that square. There are 15,729 weighted demand points representing the whole population in this region. Fig. 1 shows that the population in the studied region is highly asymmetrically distributed, with the majority living in the southeast part.

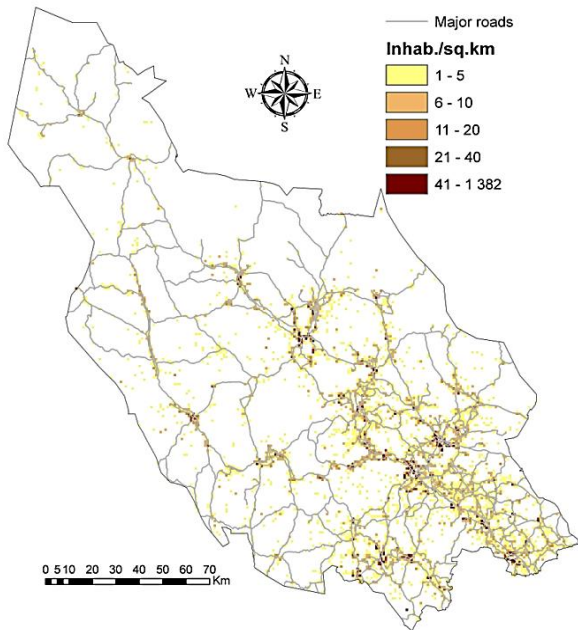


Figure 1. Map of the Dalarna region showing 250-by-250 meter squares of inhabitants

Fig. 2 shows the complete digitalized representation of the road network in Dalarna, with its hierarchical structure corresponding to 10 road levels. Fig. 2(a) shows that the main structure of the road network is composed of levels 0-5, which are European highways, national and regional roads with a total length of 5,479 km. The spatial structure of the road network of these big roads corresponds to the spatial distribution of the population, as illustrated in Fig. 1. The European highway forms the simplest road network with a limited number of candidate nodes for locating facilities. While the most complex road network includes all the roads up to the smallest level of the local and private streets, where there is a huge number of candidate nodes for locating facilities. Fig. 2(b) shows

the small roads at levels 6-9, with a total length of 33,975 km.

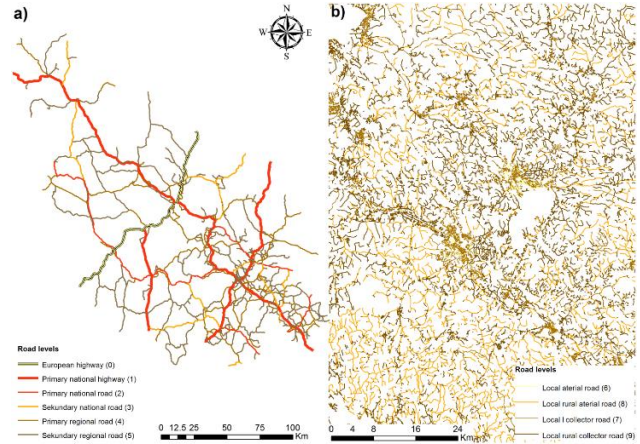


Figure 2. (a) Main structure of the road network corresponding to road levels 0-5 in Dalarna, Sweden (b) Local dense road network corresponding to road levels 6-9 in Dalarna, Sweden

Specifically, there are 1,797,939 nodes (the start and the end of the roads), and 1,964,801 road segments in the whole road network in Dalarna. Approximately, one node can be found asymmetrically distributed every 20meters on the road network. In order to have a valid measure to make comparisons between different network densities, the whole network is used to calculate all the network distances between the demand points and the candidate nodes, which can also ensure that the variation of candidate points from different network densities does not affect the distance measure. However, considering the computation complexity in solving the *NP*-hard *p*-median model, the whole network with all candidate nodes is not practical for use with most locations problems. A 500 by 500 meter grid aggregation on all density level roads was conducted for mainly two reasons: one reason is to enable all the density levels computationally feasible, and the other reason is to practically stick to the real world situation, since it is very rare to locate a facility every 20 meters.

TABLE I. NUMBER OF NODES IN DIFFERENT ROAD NETWORK COMPLEXITIES THAT CORRESPOND TO THE INCLUDED ROAD LEVELS UNDER THE GRID AGGREGATION LEVEL OF 500M BY 500M IN DALARNA

| Road Densities | Density Level               | Number of Nodes | Length (km) |
|----------------|-----------------------------|-----------------|-------------|
| 0              | European highway            | 452             | 167         |
| 0-1            | +Primary national highway   | 1,994           | 883         |
| 0-2            | +Primary national road      | 2,909           | 1,299       |
| 0-3            | +Secondary national road    | 3,926           | 1,725       |
| 0-4            | +Primary regional road      | 6,735           | 2,923       |
| 0-5            | +Secondary regional road    | 12,417          | 5,479       |
| 0-6            | +Local arterial road        | 12,552          | 5,631       |
| 0-7            | +Local collector road       | 20,718          | 10,964      |
| 0-8            | +Local rural arterial road  | 45,336          | 23,086      |
| 0-9            | +Local rural collector road | 67,020          | 39,454      |

<sup>2</sup> www.scb.se

In order to calculate the network distances, the connectivity of the nodes in the network is first checked. There are 9020 not connected and they are removed. All remaining nodes are tagged to the closest demand origin, or to the closest intersection node. In each grid, at the most, one node is kept as a potential candidate node for optimal location by applying three criteria sequentially. First, we choose the node connects with most road segments. If there are none, we then select the node at the highest density level. If there are none there, we at last pick the node which is closest to the center of the grid. Table I summarizes the numbers of nodes according to the complexity of the road network; the number of nodes increases when the complexity of the road network increases by including more levels of roads. There are 452 nodes in the simplest road network, and 67, 020 nodes in the most complex road network.

As the coordinates of the residents (demand points) do not perfectly coincide with the nodes in the road network, correspondingly, we use the nearest node in the network to represent the location of residents. The average distance between the residents' node and the nearest network node is 62 meters, which means the approximation does not introduce much difference to the final results.

The travel distance between the demand point and the nearest facility is one key variable that is in the objective function of the  $p$ -median model, therefore, the distance measure influences the solutions for optimal facility locations. Carling *et al.* ([1], [8]) investigated, empirically, the consequences of different distance measures for the optimal location of multiple service centers in rural areas. They stated that 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 reason for this is the difficulty and cost associated with collecting data on travel time. Another reason is the complication which arises in modelling the inherent variation in travel time. This paper employs network distance as the distance measure.

#### B. Settings for Sensitivity Check

Heuristics and approximation algorithms are the predominant techniques used for solving the  $p$ -median location problem, as described and explored in the recent literature ([24]-[27]). Four algorithms that work rather differently in solving the  $p$ -median model to find the solutions of optimal facility locations are chosen: CPLEX; greedy search [10]; simulated annealing [11]; and imp-Genetic algorithm [12].

If the  $p$ -median model is formulated as a 0-1 binary programming problem it can then be solved by a Mixed Integer Problem (MIP) solver by using a branch-and-cut approach. The CPLEX from IBM is a commonly used software package for solving optimization problems. Following Rebreyend *et al.* [12], some parameters of the solver have been tuned in order to adapt CPLEX to work on large problem instances, specifically, removing default computation time limits, allowing intermediate data

storage, and tuning branch-and-cut search tree strategies, according to the manual<sup>3</sup>.

The standard greedy algorithm for the  $p$ -median model was studied by Cornuejols *et al.* [28]. Resende and Werneck [29] conducted a constructive greedy algorithm to perform the most profitable move among candidates to get the best local minimum in the path. The Greedy algorithm follows the problem-solving heuristic of finding the locally optimal choice at each stage, with the hope of getting a global optimum. It always chooses the optimal choice at the current stage, rather than considering all other conditions to ensure finding a global optimal. This can be characterized as being 'short sighted', but it is easy to implement and can achieve acceptable results within a short time.

Simulated annealing (SA) is one commonly used heuristics for solving the  $p$ -median model. Murray and Church [30] proposed a basic SA algorithm for the  $p$ -median model. Levanova and Loresh [31] studied the SA heuristic that used the 1-interchange neighborhood structure. Carling *et al.* ([1], [8]), and Rebreyend *et al.* [9] used tuned SA to solve specific  $p$ -median model in a real road network context. The basic idea of SA is not only accepting all the better results in the search process, but also accepting some worse results based on certain probabilities. It is simple to implement and can provide high quality solutions to many problems. 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 employed by Rebreyend *et al.* [9]. The specific parameter settings and the implementation of SA follow Zhao *et al.* [32]. This includes their dynamic scheme to update the temperature, which allows the algorithm to have setting for efficiency and accuracy, regardless of the size of the input. After testing with various parameter settings, we found that the scheme used to increase and decrease the temperature works well to avoid the search being trapped in the local optimal for a long time, and finally provides satisfying results.

Genetic algorithms (GAs) are another commonly used heuristics that are designed based on mimicking the evolution process. New solutions are based on previous solutions in ways that are reminiscent of the interaction of genes. Most previous studies used a classical string representation, in which each chromosome is represented as a single string of length  $p$ , embedding the index of the selected facilities or nodes. Thorough treatments of GAs can be found in ([33]-[36]). Bozkaya *et al.* [37] found GAs could produce solutions that are better than exchange algorithms; however, the convergence is very slow. Alp *et al.* [38] proposed a GA which is simpler and produces good solutions faster. Following Correa *et al.* [39], Rebreyend *et al.* [12] proposed an improved genetic algorithm, called imp-GA, to solve large-scale  $p$ -median problems. The imp-GA used in this paper follows the description and settings as used by Rebreyend *et al.* [12].

<sup>3</sup>[http://www01.ibm.com/support/knowledgecenter/SSSA5P\\_12.6.1/ilog.odms.studio.help/pdf/gscplex.pdf](http://www01.ibm.com/support/knowledgecenter/SSSA5P_12.6.1/ilog.odms.studio.help/pdf/gscplex.pdf)



The facility numbers vary from 5 to 50 (with a common difference of 5) to simulate the intra-urban to inter-urban location distributions. We run CPLEX and Greedy once, due to the deterministic property of the methods. Since SA is sensitive to its starting point, we randomly select 3 different initial configurations to conduct the SA, and keep the solution with the minimum objective function value (OFV); each run contains 20,000 iterations. As for imp-GA, we set 3 runs with 100 iterations, due to its higher requirement for computation time and computer memory. Max running time is 48 hours for each  $p$  and network density level. In this paper, the CPLEX version 12.6, Linux 64 bits is used. All the programs are coded with C and compiled using GCC version 4.8.2, and they are launched under a system of Linux (Kernel 3.11-2-amd64). The computer has a memory of 32 G, CPU of Intel Core i7-3770.

IV. HELPFUL HINTS

A. Sensitivity to Complexity Levels

Fig. 3 shows how worse the optimal solutions found at each complexity level of the road network compared to the best solution found for a selection of  $p$  (results for all tested  $p$  can be found in Appendix, Figures A1-A3). Specifically, the improvement of the solution is measured by checking the difference between the OFV in the current complexity level, and the optimal OFV found at the same/other level:

$$\frac{|\text{current solution} - \text{best solution}|}{\text{current solution}} * 100\%$$

The figure reveals that the solutions found from candidate nodes on the simplest road network, namely, the European highways (level 0), are 85 to 95 percent worse than the optimal solutions that can be found. However, the figure also shows that there is a limit to how complex the road network needs to be to find the best solutions. In general, when the complexity of the

road network continues to increase after including road level 5, the solutions barely improve, and even deteriorate.

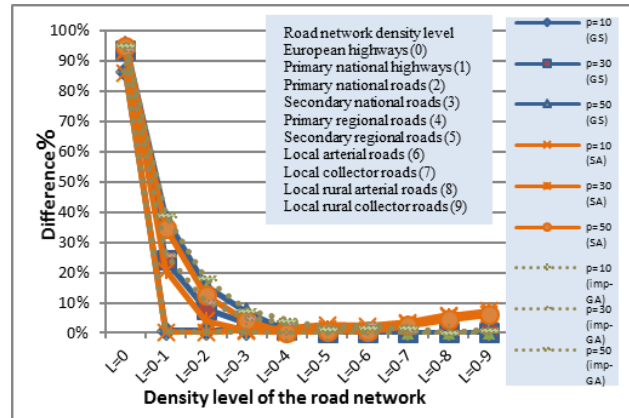


Figure 3. Variations from algorithms Greedy search, SA and imp-GA in excess distances (in percent), compared to the best solutions for an increased density level. The x-axis shows the density level. The y-axis shows the difference in percentage between the best solution and the current.

The figure reveals that the solutions found from candidate nodes on the simplest road network, namely, the European highways (level 0), are 85 to 95 percent worse than the optimal solutions that can be found. However, the figure also shows that there is a limit to how complex the road network needs to be to find the best solutions. In general, when the complexity of the road network continues to increase after including road level 5, the solutions barely improve, and even deteriorate.

B. Sensitivity to Algorithms

As is shown in Fig. 3, in general, the algorithms behave in a similar pattern of obvious improvement when the complexity increases from the simplest level 0 to levels up to, and including, 0-5 (ca. 12,500 nodes). When the complexity level increases from 0-5, the OFVs from SA deteriorate, the OFVs from the Greedy search neither improve nor deteriorate, and the OFVs from imp-GA show some very small improvements (also see Table II).

TABLE II. THE OPTIMAL OBJECTIVE FUNCTION VALUE FOUND ON DIFFERENT ROAD NETWORK DENSITIES BY TESTED ALGORITHMS FOR VARIOUS  $p$ .

|               |                            |   |                  |                 |                 |                 |                 |                 |                 |                 |                 |
|---------------|----------------------------|---|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <b>CPLEX</b>  | Included complexity level  | 0   |                  |                 |                 |                 |                 |                 |                 |                 |                 |
|               | Optimal solution in meters | 81681 (The software aborts before completion on our computer with 32 Gb of memory). |                  |                 |                 |                 |                 |                 |                 |                 |                 |
| <b>Imp-GA</b> | Included complexity level  | 0-5   | 0-7              | 0-8             | 0-5             | 0-6             | 0-8             | 0-9             | 0-8             | 0-8             | 0-8             |
|               | Optimal solution in meters | <b>19624.095</b>  | <b>11075.455</b> | <b>8278.562</b> | <b>6631.992</b> | <b>5715.777</b> | <b>4998.492</b> | <b>4600.623</b> | <b>4250.913</b> | <b>3960.126</b> | <b>3738.607</b> |
| <b>SA</b>     | Included complexity level  | 0-2   | 0-2              | 0-4             | 0-6             | 0-5             | 0-4             | 0-6             | 0-4             | 0-4             | 0-4             |
|               | Optimal solution in meters | 19715.020   | 11259.282        | 8625.458        | 7094.882        | 6155.251        | 5465.033        | 5013.738        | 4611.287        | 4306.700        | 4066.063        |
| <b>Greedy</b> | Included complexity level  | 0-5   | 0-5              | 0-5             | 0-5             | 0-5             | 0-5             | 0-5             | 0-5             | 0-5             | 0-5             |
|               | Optimal solution in meters | 19685.104   | 11134.182        | 8316.884        | 6698.967        | 5804.399        | 5122.226        | 4708.024        | 4349.791        | 4049.786        | 3796.301        |
| <b>p</b>      |                            | <b>5</b>  | <b>10</b>        | <b>15</b>       | <b>20</b>       | <b>25</b>       | <b>30</b>       | <b>35</b>       | <b>40</b>       | <b>45</b>       | <b>50</b>       |

Table II further shows how the best OFVs derived from the four algorithms vary among these algorithms. Note that the CPLEX algorithm that gives the exact solutions fails to produce a result higher than density level 0. The reason for this is that the location problem addressed in this study is too complex, even when the representation of the road network is at the simplest level (level 0). Imp-GA gives the best solutions. However, for the same facility number, the improvements of the OFVs are rather small compared to Greedy and SA, and the largest improvement is only 2.78%, i.e. approximately 122 meters shorter in network travel distance.

Table II also shows that the algorithms show a different requirement of network complexity for finding better solutions by including the road levels. However, the differences between the Greedy and imp-GA OFVs are just 1-2 percent. For  $p=20$ , or larger, the SA gives an OFV that is 8 to 9 percent worse than the OFV from the imp-GA. Whereas, the computation time and computation effort of imp-GA become tremendously large, so that the results cannot be derived for the most complex of the road networks within the time threshold of 48 hours, which is not the case for Greedy and SA

C. Sensitivity to Facility Numbers

Based on Fig. 3 and Table II, the changes of facility numbers do not introduce obvious changes either in the complexity level or among the algorithms. Fig. 4 visualizes the spatial distribution of the optimal locations found by imp-GA.

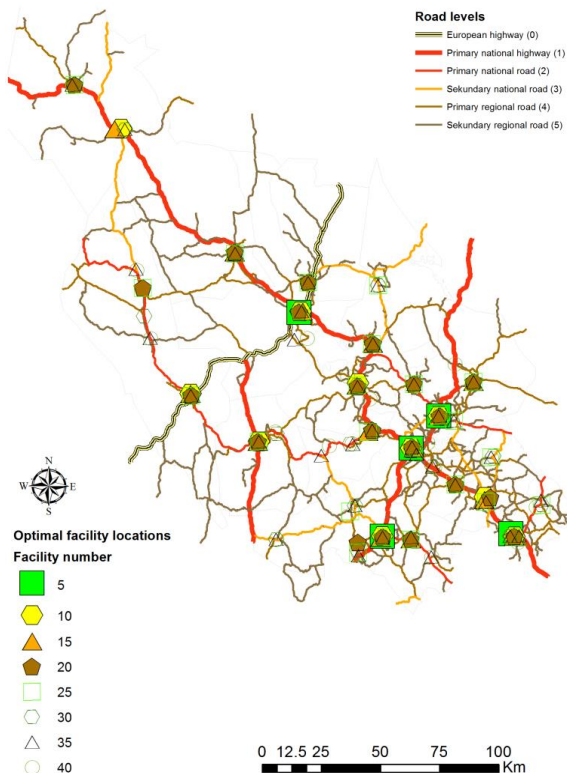


Figure 4. Distribution of optimal locations (p from 5 to 40) found by imp-GA on the study area Dalarna

The road network has hierarchical road levels and an asymmetrical structure, where the European highway

forms the simplest road network with 452 candidate nodes, while the most complex road network includes all the roads up to the smallest level of the local and private streets, with 67, 020 candidate nodes.

In general, the optimal locations found when  $p$  is low remain as optimal locations when  $p$  increases. The locations are mostly on the major roads in the road network and correspond very well with the complexity structure, as shown in Fig. 2(a). This further indicates that the increase in complexity of the road network above level 0-5 does have a limited influence on the spatial distribution of the optimal facility locations. To make the location-allocation practical in transportation and land-use planning, a location that decreases the travel distance a few hundred meters is not appealing, compared to the risk of losing the attraction of customers due to poor accessibility on a denser road network, especially when the road network is asymmetrically distributed.

V. CONCLUDING DISCUSSION

This paper aims to investigate how the changes of road network complexity influence the optimal facility locations by applying the widely used p-median model. In addition, to provide further insights on computation complexity and location problems from intra-urban to inter-urban, a detailed sensitivity analysis of four algorithms and various facility numbers is conducted on a specific case study, namely, Dalarna province, Sweden.

The road network has hierarchical road levels and an asymmetrical structure, where the European highway forms the simplest road network with 452 candidate nodes, while the most complex road network includes all the roads up to the smallest level of the local and private streets, with 67, 020 candidate nodes.

The main result indicates that there is a limit to how complex the road network needs to be to find the best solutions. When the complexity of the road network continues to increase after a certain level the solutions barely improve, but even deteriorate. We found that the choice of algorithms and facility number  $p$  has little to no impact on the results, although due to differences in the mechanics of the algorithm, there are some variations among the algorithms. The most complex level of road network (including local streets) is not preferable, especially when the number of the facility is less than 20. The spatial distributions of the optimal locations show that optimal locations found for smaller  $p$  remain as part of the optimal locations when  $p$  increases.

The specific complexity level that is found efficient here might be case-specific, and additional empirical analysis on other cases is needed. In the p-median model, people are assumed to always choose the closest facility and neglect the multiple purpose travel, or the heterogeneous preferences on facilities. Therefore, using the gravity p-median model [40] and other algorithms for MinMax location problems on road network ([41], [42]) would provide more insight into the impact of road network complexity on optimal facility locations for aiding the decision-making in location and transportation planning.

APPENDIX

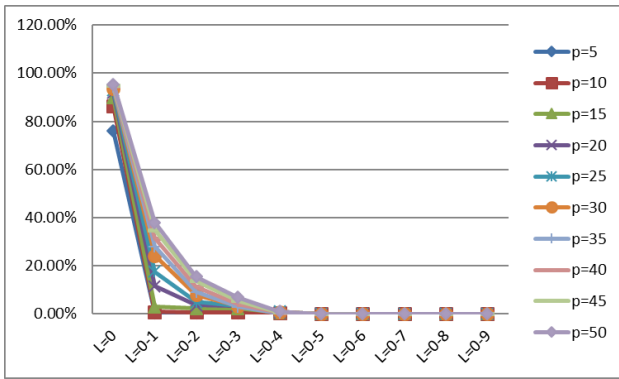


Figure A1. Variations from Greedy in excess distances (in percent) compared to the best solutions for an increased density level. The x-axis shows the density level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to  $\left(\frac{\text{current solution} - \text{best solution}}{\text{current solution}}\right) * 100\%$ .

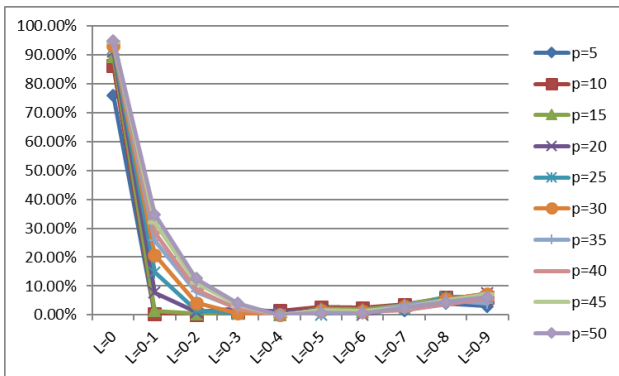


Figure A2. Variations from SA in excess distances (in percent) compared to the best solutions for an increased density level. The x-axis shows the density level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to  $\left(\frac{\text{current solution} - \text{best solution}}{\text{current solution}}\right) * 100\%$ .

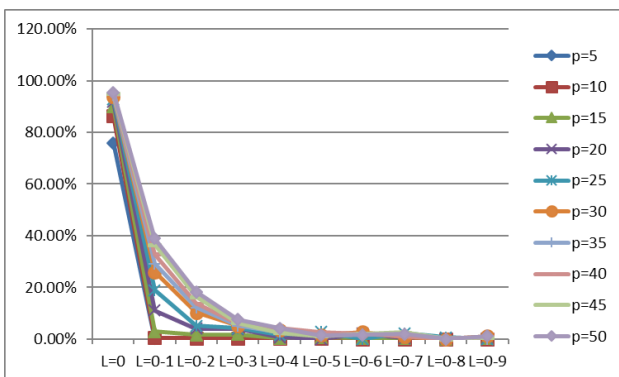


Figure A3. Variations from imp-GA in excess distances (in percent) compared to the best solutions for an increased density level. The x-axis shows the density level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to  $\left(\frac{\text{current solution} - \text{best solution}}{\text{current solution}}\right) * 100\%$ .

REFERENCES

[1] K. Carling, M. Han, and J. Håkansson, "Does euclidean distance work well when the p-median model is applied in rural areas?" *Annals of Operations Research*, vol. 201, no. 1, pp. 83-97, 2012.

[2] S. Kelley, "AFV refueling stations and the complexity of freeway intersections: The scale dependency of network representation," *International Journal of Geographical Information Science*, vol. 31, no. 2, pp. 346-363, 2017.

[3] Z. Tian, L. Jia, H. Dong, F. Su, and Z. Zhang, "Analysis of urban road traffic network based on complex network," *Procedia Engineering*, vol. 137, pp. 537-546, 2016.

[4] S. Melkote and M. S. Daskin, "An integrated model of facility location and transportation network design," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 6, pp. 515-538, 2001.

[5] D. Peeters and I. Thomas, "The effect of spatial structure on p-median results," *Transportation Science*, vol. 29, no. 4, pp. 366-373, 1995.

[6] D. Schilling, K. Rosing, and C. Reville, "Network distance characteristics that affect computational effort in p-median location problems," *European Journal of Operational Research*, vol. 127, no. 3, pp. 525-536, 2000.

[7] R. Francis, T. Lowe, M. Rayco, and A. Tamir, "Aggregation error for location models: Survey and analysis," *Annals of Operations Research*, vol. 167, no. 1, pp. 171-208, 2009.

[8] K. Carling, M. Han, J. Håkansson, and P. Rebreyend, "Distance measure and the p-median model in rural areas," *Annals of Operations Research*, vol. 226, no. 1, pp. 89-99, 2015.

[9] P. Rebreyend, M. Han, and J. Håkansson, "How does different algorithm work when applied on the different road networks when optimal location is searched for in rural areas?" in *Proc. Web Information Systems Engineering-WISE 2013 Workshops*, Nanjing, China: Springer Berlin Heidelberg, 2013, pp. 284-291.

[10] A. A. Kuehn and M. J. Hamburger, "A heuristic program for locating warehouses," *Management Science*, vol. 9, no. 4, pp. 643-666, 1963.

[11] Al-Khedhairi, "Simulated annealing metaheuristic for solving p-median model," *International Journal of Contemporary Mathematical Sciences*, vol. 3, no. 28, pp. 1357-1365, 2008.

[12] P. Rebreyend, L. Lemarchand, and R. Euler, "A computational comparison of different algorithms for very large p-median models," *Evolutionary Computation in Combinatorial Optimization*, pp. 13-24, 2015.

[13] S. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of graph," *Operations Research*, vol. 12, no. 3, pp. 450-459, 1964.

[14] R. Farahani, N. Asgari, N. Heidari, M. Hosseini, and M. Goh, "Covering problems in facility location: A review," *Computers and Industrial Engineering*, vol. 62, no. 1, pp. 368-407, 2012.

[15] O. Kariv and S. Hakimi, "An algorithmic approach to network location problems. II: The p-medians," *SIAM Journal on Applied Mathematics*, vol. 37, no. 3, pp. 539-560, 1979.

[16] Al-Khedhairi, "Simulated annealing metaheuristic for solving p-median model," *International Journal of Contemporary Mathematical Sciences*, vol. 3, no. 28, pp. 1357-1365, 2008.

[17] W. Yu, T. Ai, and S. Shao, "The analysis and delimitation of central business district using network kernel density estimation," *Journal of Transport Geography*, vol. 45, pp. 32-47, 2015.

[18] Z. Xie and J. Yan, "Kernel density estimation of traffic accidents in a network space," *Computers, Environment and Urban Systems*, vol. 35, no. 5, pp. 396-406, 2008.

[19] S. Shiode, "Street-level spatial scan statistic and STAC for analysing street crime concentrations," *Transactions in GIS*, vol. 15, no. 3, pp. 365-383, 2011.

[20] T. Ai, W. Yu, and Y. He, "Generation of constrained network Voronoi diagram using linear tessellation and expansion method," *Computers, Environment and Urban Systems*, vol. 51, pp. 83-96, 2015.

[21] D. Peeters, J. F. Thisse, and I. Thomas, "Transportation networks and the location of human activities," *Geographical Analysis*, vol. 30, no. 4, pp. 355-371, 1998.

[22] J. F. Bigotte, D. Krass, A. P. Antunes, and O. Berman, "Integrated modeling of urban hierarchy and transportation network planning," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 7, pp. 506-522, 2010.

[23] R. Rahmani and A. Ghaderi, "A combined facility location and network design problem with multi-type of capacitated links," *Applied Mathematical Modelling*, vol. 37, no. 9, pp. 6400-6414, 2013.

- [24] B. C. Tansel, R. L. Francis, and T. J. Lowe, "State of the art—location on networks: A survey. Part I: the p-center and p-median models," *Management Science*, vol. 29, no. 4, pp. 482-497, 1983.
- [25] J. Reese, "Solution methods for the p-median model: An annotated bibliography," *Networks*, vol. 48, pp. 125-142, 2006.
- [26] N. Mladenović, J. Brimberg, P. Hansen, and J. A. Moreno-Pérez, "The p-median model: A survey of metaheuristic approaches," *European Journal of Operational Research*, vol. 179, no. 3, pp. 927-939, 2007.
- [27] M. J. Varnamkhasti, "Overview of the algorithms for solving the p-median facility location problems," *Advanced Studies in Biology*, vol. 4, no. 2, pp. 49-55, 2012.
- [28] G. Cornuejols, M. L. Fisher, and G. L. Nemhauser, "Exceptional paper-location of bank accounts to optimize float: An analytic study of exact and approximate algorithms," *Management Science*, vol. 23, no.8, pp. 789-810, 1977.
- [29] M. G. Resende and R. F. Werneck, "A hybrid heuristic for the p-median model," *Journal of Heuristics*, vol. 10, no. 1, pp. 59-88, 2004.
- [30] A. T. Murray and R. L. Church, "Applying simulated annealing to location-planning models," *Journal of Heuristics*, vol. 2, no. 1, pp. 31-53, 1996.
- [31] T. Y. Levanova and M. A. Loresh, "Algorithms of ant system and simulated annealing for the p-median model," *Automation and Remote Control*, vol. 65, no. 3, pp. 431-438, 2004.
- [32] X. Zhao, K. Carling, Z. Dan, and J. Håkansson, *Network Density and the P-median Solution*, Working Papers in Transport, Tourism, Information Technology and Microdata Analysis, Dalarna University, Sweden, 2013.
- [33] L. Davis, *Genetic Algorithms and Simulated Annealing*, Morgan Kaufman Publishers, Inc., Los Altos, CA, 1987.
- [34] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 1991.
- [35] C. R. Reeves, "Using Genetic Algorithms with Small Populations," in *Proc. 5th International Conference on Genetic Algorithms*, S. Forrest, ed., Morgan Kaufmann, San Mateo, CA, 1993, pp. 92-99.
- [36] K. A. Dowsland, "Genetic algorithms-a tool for OR?" *Journal of the Operational Research Society*, pp. 550-561, 1996.
- [37] B. Bozkaya, J. Zhang, and E. Erkut, "An efficient genetic algorithm for the p-median model," *Facility Location: Applications and Theory*, pp. 179-205, 2002.
- [38] O. Alp, E. Erkut, and Z. Drezner, "An efficient genetic algorithm for the p-median model," *Annals of Operations Research*, vol. 122, no. 1-4, pp. 21-42, 2003.
- [39] E. S. Correa, M. T. Steiner, A. A. Freitas, and C. Carnieri, "A genetic algorithm for the p-median model," in *Proc. Genetic and Evolutionary Computation Conference*, 1993, pp. 1268-1275.
- [40] T. Drezner and Z. Drezner, "The gravity p-median model," *European Journal of Operational Research*, vol. 179, no. 3, pp. 1239-1251, 2007.
- [41] F. Barahona and R. Anbil, "The volume algorithm: Producing primal solutions with," *Mathematical Programming*, vol. 87, no. 3, pp. 385-399, 2000.
- [42] R. Liu, A. W. C. Fu, Z. Chen, S. Huang, and Y. Liu, "Finding multiple new optimal locations in a road network," in *Proc. 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, October 2016, pp. 36-47.

**Xiaoyun Zhao** born in 1987 in China and got PhD degree in Microdata analysis at Dalarna University, Sweden in 2017 focus on sustainable residential planning and CO<sub>2</sub> emissions.

She now works as a post doc researcher at integrated transport research lab of KTH Royal Institute of Technology with a focus of system effects of self-driving vehicles and a lecturer at Dalarna University, both in Sweden. She is one of the author of book Working with Swedes.

**Pascal Rebreyend** is a researcher born in France in 1972. He got a PhD in computer science from Ecole Normale Supérieure, Lyon, France, in 1998 about hybrid genetic algorithms in combinatorial optimization.

After his PhD, he work as Teaching Assistant University of Orsay, France and Université of Saint-Etienne France. He worked then as Engineer in Grenoble, in charge of testing and validating C compilers for an embedded processor. He moved in 2001 to Dalarna University, Sweden and worked as Senior Lecturer.

**Johan Håkansson** born in 1966 in Sweden and got his Ph. D. in human geography in Umeå University the year 2000.

He received the associate professorship in human geography in 2013 in Dalarna University. Associate professor Håkansson is the program leader in Microdata analysis and actively work in the Ph.D. supervisor steering committee.