Studying Solutions of the p-Median Problem for the Location of Public Bike Stations

Christian Cintrano[1], Francisco Chicano[1], Thomas Stützle[2], and Enrique Alba[1]

[1] University of Málaga Andalucía Tech,, E.T.S. Ingeniería Informática
Bulevar Louis Pasteur 35, 29071 Málaga, Spain
{cintrano,chicano,eat}@lcc.uma.es
[2] Universite Libre de Bruxelles, CoDE, IRIDIA
Av. F. Roosevelt 50 B-1050 Brussels, Belgium
stuetzle@ulb.ac.be

Abstract. The use of bicycles as a means of transport is becoming more and more popular today, especially in urban areas, to avoid the disadvantages of individual car traffic. In fact, city managers react to this trend and actively promote the use of bicycles by providing a network of bicycles for public use and stations where they can be stored. Establishing such a network involves the task of finding best locations for stations, which is, however, not a trivial task. In this work, we examine models to determine the best location of bike stations so that citizens will travel the shortest distance possible to one of them. Based on real data from the city of Malaga, we formulate our problem as a p-median problem and solve it with a variable neighborhood search algorithm that was automatically configured with irace. We compare the locations proposed by the algorithm with the real ones used currently by the city council. We also study where new locations should be placed if the network grows.

Keywords: Bike station location · p-Median problem · Variable neighborhood search.

1 Introduction

Driving a vehicle through the city is an increasingly difficult and annoying task. A large number of traffic jams at different times of the day, the rising cost of fuel, and the rising level of pollution in cities are some of the biggest problems for citizens. For all these reasons, more and more citizens are looking for alternative and sustainable ways to move around the city. Bicycles are a good way to get around: they are clean and environmentally friendly, beneficial to the health, and help to avoid getting trapped in traffic jams. As a result, an increasing number of people use bicycles as their main means of transport in the city. Municipalities have become aware of this trend and try to promote their use by providing the necessary infrastructure, such as areas to place them and bike lanes. In addition, numerous initiatives have been taken by local authorities and private companies to promote the so-called bike-sharing. To manage the distribution of bicycles,

stations are usually set up where bicycles can be picked up/dropped off. But finding the best location for those facilities is not a trivial job.

In this work, we tackle the public bike stations location problem (PBSLP). To do so, we formulate the problem as a classical problem of localization, namely the p-median problem. The p-median problem tries to identify, given a set of locations and customers and distances between locations and customers, a subset of locations of size p such that the total distance of travel of customers to the closest location is minimized. In other words, the goal is to identify locations so that the average distance of customers to bike stations is minimized. While we are not the first to consider this problem to a possible model for the PBSLP [17], we study here different variants of the problem including different distance metrics. This classic formulation is accompanied by real data from the city of Malaga, Spain, which allows us to test our proposal and variants of the p-median problem as close as possible to the reality of the city. Finally, an advantage of this modeling is that the p-median problem is well-studied and many efficient and effective algorithms are available for it. We further enhance the performance of some popular p-median algorithms by an automatic algorithm design process.

The rest of this article is organized as follows: Section 2 presents the formulation of the problem. Section 3 describes the selected optimization algorithm. Section 4 analyzes the main results of our work using as a real scenario the city of Malaga. We discuss related work in Section 5 and conclude in Section 6.

2   The p-median problem

The p-median problem is one of the most-studied NP-hard discrete location problems [6, 14]. The problem can be formulated as follows. Given a set of customers N and a set of possible facility locations F, the p-median problem asks to allocate p facilities to the set of available locations F while minimizing the weighted sum of the distances between the customers and their closest facility. Formally, the optimization problem is defined as:

$$\min \sum_{i=1}^{|N|} w_i \min_{j \in L} d_{ij}, \tag{1}$$

where $L \subseteq F$, $|L| = p$, w_i is the weight of customer i, and d_{ij} is the distance between customer i and facility j. If we have $w_i = 1$, $i = 1, \ldots, |N|$, that is, all weights are one, we have the unweighted version of the problem.

In our study the customers are the citizens and the facilities are the bicycle stations that can be placed on different street segment of the city. We selected this formulation of the problem for two reasons: (i) the model is easy to understand and implement and (ii) the p-median problem is a classical location problem that has been well-studied in the scientific literature [6]. From the PBSLP perspective, the p-median may serve as a good proxy to identify interesting locations and relevant possible distributions of the bicycle stations across an urban area.

While there are other formulations of the problem [2, 12, 10], these may in part require various types of information. Differently, the p-median problem requires rather little information (only the distance matrix and possibly the weights). Although this may seem like a limitation, it is relatively simple to add additional information, either by pre-processing the weights or distances, (e.g., by considering the slopes of the streets) or by adding terms in the formulation itself (e.g., by adding capacity information related to the number of bikes in each bicycle parking site). Hence, in the present work we define a baseline for future works that will take these advanced characteristics into account.

3   Algorithm

The algorithm, that we use in this article, is based on the variable neighborhood search (VNS) algorithm for the p-median problem presented by Mladenović and Hansen [15]. It was selected as the base for our development because the VNS algorithms and variants of it have been used in localization problems, and, in particular, for the p-median problem reaching very good performance [1, 16, 18].

While there are a number of VNS variants for the p-median problem, we have developed a component-wise implementation of such algorithms. In particular, we use algorithmic components that have been used for the original, basic VNS algorithm [15] (OVNS), and two more recent variants proposed in [7] (BVNS and DVNS). In order to find an improved version of the VNS, we use an automatic algorithm configuration stage through the use of the iterated racing procedure implemented by the irace package [13]. We used a budget (number of configurations) of 5000 and, as training instances, we chose instances of the TSPLIB[3] library. The locations of the cities correspond to the locations of the customers and the possible facilities. As in the practical problem we tackle the number of p with small values, we used here values of $p \in \{10, 20, 30\}$ for each instance of the set of training instances. For the configuration process we considered the following components together with associated numerical parameters:

Neighborhood model. We used two models to select the (k) neighbors: nearest points (NEAR) and the division of the space into four quadrants from the point and selecting (if possible) an equal number of neighbors in each quadrant (QUAD).

Local search. We apply a local search procedure to the initial solution (localsearch1) and in each iteration of the algorithm (localsearch2). The possible local search algorithms are: Fast Interchange [20], IALT [8] (Laux parameter), and IMP [7].

Shake. We use two different solution modifications: either we exchange n facilities with randomly chosen other ones in the neighborhood (NEIGHBOR) or we select then randomly in the whole set of facilities (RAND).

Initial solution. We have four procedures to generate the initial solution: random (RAND), START [7], the best of 100 random solutions (100RAND), and the best of performing IMP to 100 different random solutions (100IMP).

---

[3] TSPLIB instances: http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/tsp/

Table 1. Parameters for the automatic configuration process. Given are the parameter, its type, its domain of possible values, the condition under which the parameter is relevant (conditional parameters) and the value the parameter takes in the automatically obtained configuration.

Parameter	Type[a]	Range	Condition	Configured
algorithm	c	$(OVNS, DVNS, BVNS)$		$DVNS$
localsearch1	c	$(FI, IALT, IMP)$		IMP
Laux1	i	$(1, 50)$	$localsearch1 = "IALT"$	—
localsearch2	c	$(FI, IALT, IMP)$	$algorithm \in ("DVNS", "BVNS")$	FI
Laux2	i	$(1, 50)$	$localsearch2 = "IALT"$	—
generation	c	$(RAND, START, 100RAND, 100IMP)$		100IMP
theta	r	$(0,1)$	$generation = "START"$	—
shake	c	$(RAND, NEIGHBOR)$		$NEIGHBOR$
neighborhood	c	$(NEAR, QUAD)$		$QUAD$
k	i	$(1, 40)$		40
Kmayus	i	$(10, 100)$	$algorithm \in ("DVNS", "BVNS")$	92
m	r	$(0, 1)$	$algorithm = "DVNS"$	0.12
lambda	r	$(1, 5)$	$algorithm = "DVNS"$	4.30

[a] c: categorical, i: integer, r: real;

Table 1 presents the parameters information used by irace. The last column shows the best values found by irace. These values form the parameters of the VNS algorithm used in the experiment. While we leave a detailed evaluation of the configured VNS algorithm to previously proposed VNS algorithms to an extended version of the paper, in what follows we focus on the usage of the configured algorithm to study the location of public bike stations in Malaga.

4   Experimental study

As mentioned, our study focuses on the usage of the p-median problem to suggest possible solutions that may be implemented in the real world to tackle the PB-SAP. For this reason, the use of actual data is really important in this work. We have decided to use the city of Malaga as a study object as it has a wide variety of open data (see municipality open data website: http://datosabiertos.malaga.eu/) and has already a functioning shared bikes system with 23 public accessible stations, which will allow us to compare our proposed solutions by the p-median problem to the actual situation.

The p-median problem needs two sets of points: customers and facilities. The customers are the citizens. As positions where these are located, we have chosen the centers of the different neighborhoods (363) of the city (we have excluded six neighborhoods at the outskirts of Malaga) and the weights are the real number of inhabitants in each of the neighborhoods. All possible street segments of the city have been chosen as possible locations for the facilities, that is, the bike stations. This makes a total of 33,550 possible locations. Figure 1 shows the layout of the customers concerned and the bike stations in the city, where the red points correspond to the centers of the neighborhoods and the clue points to the current location of the 23 public bike stations. We should note there that the ratio between the number of clients and facilities we have in this instance is not common in the existing studies in the p-median problem; hence, further studies with instances with such ratios may be an interesting direction for future work.

We next present the main results of our case study and compare them with the real scenario. We used real data and the tuned VNS to optimize each in-
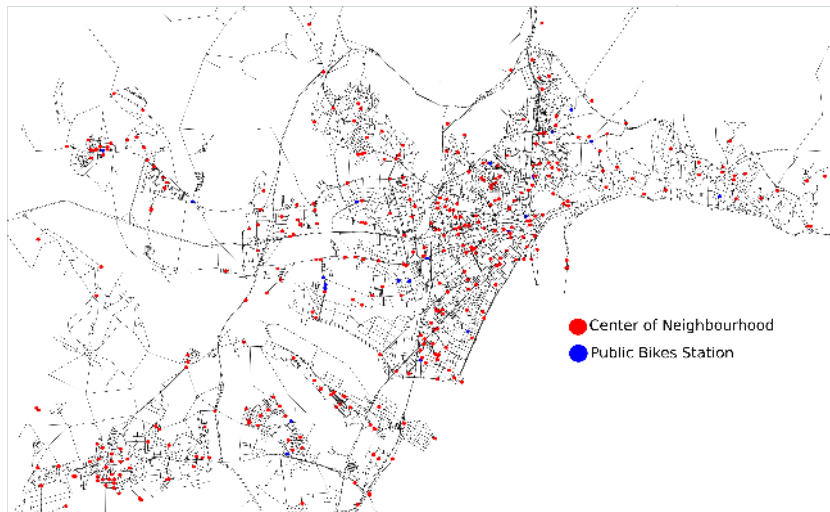
Fig. 1. Public bike stations in the city of Malaga (blue points) and the center of each neighborhood (red points).

stance. The algorithm was running 60 CPU seconds and reports the best solution found in each of the trials. The implementation was done in C++, compiled with flag -O3. The computation platform used is a desktop computer with one Intel i5-4460 processor running at 3.2 GHz, and 8 GB memory. We have carried out 100 runs of the algorithm to statistically compare the results obtained.

4.1 Comparison with real scenario

In our study we have taken into account different levels of realism to analyze its impact on the proposed solutions. We have used two types of distance between customers and stations: straight-line Euclidean distance and the shortest path through the city streets (calculated using the Dijkstra's algorithm), we refer to this last one as real distance. We have also taken into account the population density in each neighborhood by weighing the problem using the number of citizens and compare it to how solutions would look if these weights were not taken into account. This gives us four scenarios to study, corresponding to weighted or unweighted customers locations with Euclidean or real distances. We used the same number of stations that are currently in use in Malaga, that is, $p = 23$. Figure 2 shows the empirical cumulative distribution of the percentage of improvement in the objective function value of each scenario when compared to the result of calculating the objective function for the actual location of the 23 bike stations in Malaga (evaluated according to the corresponding scenario).

The versions without weights show larger improvements, with a median of 68% (Euclidean distance) and 62% (real distance), than the versions with weights, where the improvement was less: 56% (Euclidean distance) and 49% (real distance). This results can be explained because by not taking the population size of each neighborhood into account, the stations become spread across
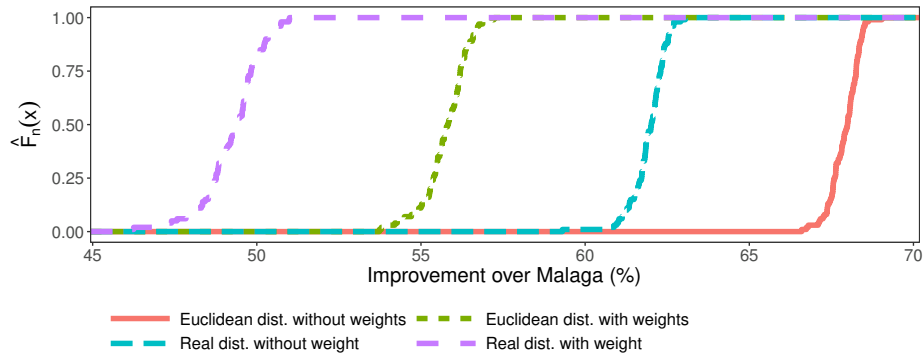
Fig. 2. Empirical cumulative distribution of the improvement of our solutions in each scenario, compared to current real location of bike stations in Malaga. For each scenario the current solution of Malaga is evaluated with the corresponding objective function.

the city, reaching more areas closer to the borders, reducing the distance to each neighborhood. However, the population size should be taken into account, as has apparently been done in the current solution implemented in Malaga, where the stations are concentrated in the central part of the city, which is also the most densely populated area. Even so, the p-median versions with weights obtain substantial improvements of about 50% improvement over the base scenario.

After analyzing the quality of the results, we can see in Figure 3 the geographical distribution in each instance of the problem. As we expected, the weighted versions put more stations in the central area of the city. However, all of them offer good coverage of the main neighborhoods, so that each citizen has a reasonably nearby station to use the service.

Finally, bike-sharing systems should try to minimize the distance users need to travel to their nearest station. Therefore, for each solution found, we calculated the average distance that the citizens must travel to their nearest station. Table 2 shows this distance information for each scenario. In general, our solutions reduce the city council's solution by 500 meters (on average per person even in the worst case). As expected, the real distance with weights obtained the best results when taking this distance into account during the optimization process. It is interesting to note that using weights in the Euclidean distance case reduces the average distance walked more than when changing the distance

Table 2. Distance traveled per inhabitant to the nearest station in the different scenarios, evaluated as real weighted distance. The minimum values are marked in bold.

Scenario	Distance (m)			
	Min	Max	Median	Mean
D. Euclidean without weights	877.3	984.2	933.3	933.1
D. Euclidean with weights	761.4	865.5	802.2	802.4
D. Real without weights	853.7	951.7	902.9	902.3
D. Real with weights	**728.2**	**801.2**	**751.3**	**754.1**
Malaga	1485.9			

Euclidean dist. without weights

Euclidean dist. with weights

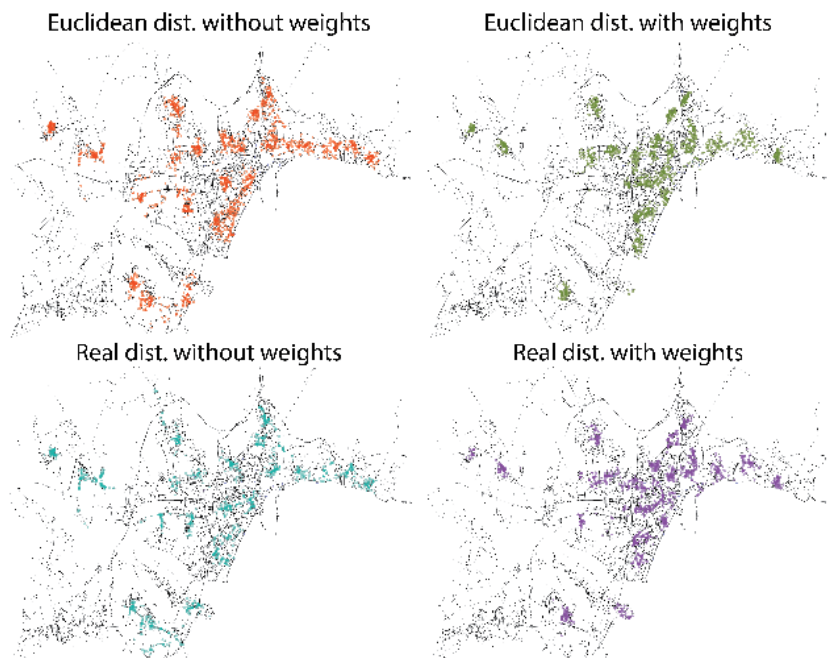Real dist. without weights

Real dist. with weights

Fig. 3. Station location of each solution and instance.

computation from Euclidean to real distances. This tells us that using demographic information is more important than the type of distance used, if one wants to improve the quality of service to the user.

4.2   Increasing number of stations

As it is is often costly to make changes to urban infrastructure, it is not always possible to move the city infrastructure. So, instead of that, we will study now the optimal locations if new stations are to be installed in the city. Using the most realistic scenario, weights and real distances, we have considered different values of p, while keeping the 23 stations in the city of Malaga as fixed in each solution. As a result, in each of the following experiments, a number of $p - 23$ bike stations is added to the current solution already existing. We have performed 100 iterations of the algorithm for $p \in \{30, 35, \ldots, 50, 60, \ldots, 100\}$. Table 3 summarizes the results obtained. With only seven additional stations, one already may obtain a substantial improvement of 43% when compared to the actual situation in Malaga. More additional stations lead obviously to additional improvements, with a steep decrease in average distances to travel for small increments and to diminishing returns when adding a large number of stations. The traveled distance is also plotted in Figure 4. The slope of this curve is quite steep up to $p = 50$ stations and smoother thereafter. As a summary, this analysis indicates that increasing the stations in the city considering optimized locations could greatly improve the quality of service to citizens.

Table 3. Walking distance for different numbers of additional stations; the number of additional stations is $p - 23$, where p is the number given in the first column.

Scenario	Distance (m)			
	Min	Max	Median	Mean
Malaga (23)		1485.9		
30	830.87	860.92	839.93	840.82
35	732.89	761.39	744.62	745.47
40	663.80	696.49	672.87	674.19
45	602.65	637.48	615.93	616.75
50	557.29	599.86	570.82	571.52
60	485.66	523.55	499.93	500.35
70	426.08	473.54	447.19	446.74
80	387.16	417.81	400.71	400.51
90	351.08	384.28	364.40	364.45
100	316.26	351.99	332.33	332.77

Comparing this data with our more realistic scenario, real distance with weights, we see that the average distance traveled (751.3) is even better than increasing Malaga's stations to 30 (being even better in some cases to $p = 35$). The power of VNS to solve this problem underscores the usefulness of algorithms to solve real problems and to consider optimization aspects upfront.

5   Related Work

In bicycle sharing systems there are multiple problems to address such as predicting the filling of stations [19], the location-allocation of bicycles [3, 5], routes for users or the transfer of bicycles [9], etc. There are also complete solutions that take into account multiple aspects of bike sharing systems [11]. However, they are very complex solutions that require large amounts of information that are not always available.

If we focus on the optimal location of the stations we find solutions such as those proposed in [2]. In this paper, the authors use real data and machine learning techniques to find the best places to position the stations. A similar approach to the above is used in [12], where they use New York City, which has a large network of bike stations. In [17] a comparison between two models for the station location problem is presented. The authors compare the p-median and the MCLP. However, they do not give details about the algorithm that is used. In [4] the authors try to find the best locations so as to ensure the availability to collect/deposit the bicycles, taking into account the demand (at peak times) and the possible routes between stations made by users. These works involve a demand on the use of the bike-sharing system. However, they do not take into account potential users who could use the system if it were closer to them.

Our work uses demographic data to bring the system closer to all users (which has not been taken into account in other work). In addition, existing solutions use a custom formulation, making it difficult to compare them with other state-of-the-art solutions. We formulate the problem of locating bike stations as a p-median which allows us to use a solid base for our studies, as well as to enrich ourselves with all the advances made in the p-median problem.
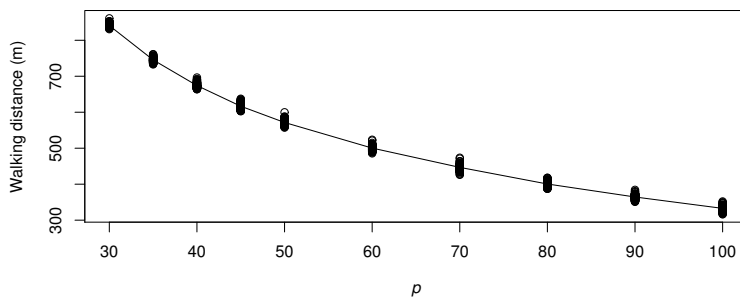
Fig. 4. Traveled distance for each new value of p.

6   Conclusions

In this work we have modeled the public bike stations location problem as a classic problem of location, the p-median problem. To solve it we used a VNS algorithm that was automatically configured with irace. We have studied different scenarios: Euclidean and real distances (calculated by Dijkstra's algorithm), and whether or not to consider population densities. As might be expected, the most realistic scenario (real distances and weights) reported the best results, and all configurations performed much better than the Malaga public bike-sharing system. Finally, we have studied where ideally locations should be added if the current public bike system is extended by new stations. Already with few additional stations, we have obtained substantial reductions in the average distance of customers to bike stations; overall the improvements have been in the range of 43-77% with respect to the current situation in Malaga.

As future work we want to test our model in bigger cities like Madrid or New York, try other heuristic algorithms and integer linear programming solvers, as CPLEX; and variants of the p-median problem such as minimizing the distance to customers while maximizing the distance between facilities or the capacitated version. It is also interesting include more realistic data such as the bike trips, cycle lanes, traffic patterns, point of interest in the city (schools, hospitals, museums, etc), or stops of other public transport, to promote multimodal transport.

References

1. Avella, P., Boccia, M., Salerno, S., Vasilyev, I.: An aggregation heuristic for large scale p-median problem. Computers & Operations Research **39**(7), 1625–1632 (2012)
2. Chen, L., Zhang, D., Pan, G., Ma, X., Yang, D., Kushlev, K., Zhang, W., Li, S.: Bike Sharing Station Placement Leveraging Heterogeneous Urban Open Data. In:

Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15. pp. 571–575. ACM Press, NY, USA (2015)

3. Chen, Q., Liu, M., Liu, X.: Bike Fleet Allocation Models for Repositioning in Bike-Sharing Systems **10**(1), 19–29 (2018)
4. Chen, Q., Sun, T.: A model for the layout of bike stations in public bike-sharing systems. Journal of Advanced Transportation **49**(8), 884–900 (dec 2015)
5. Chira, C., Sedano, J., Villar, J.R., Cámara, M., Corchado, E.: Urban bicycles renting systems: Modelling and optimization using nature-inspired search methods. Neurocomputing **135**, 98–106 (jul 2014)
6. Dantrakul, S., Likasiri, C., Pongvuthithum, R.: Applied p-median and p-center algorithms for facility location problems. Expert Systems with Applications **41**(8), 3596–3604 (jun 2014)
7. Drezner, Z., Brimberg, J., Mladenović, N., Salhi, S.: New heuristic algorithms for solving the planar p-median problem. Computers & Operations Research **62**, 296–304 (oct 2015)
8. Drezner, Z., Brimberg, J., Mladenović, N., Salhi, S.: New local searches for solving the multi-source Weber problem. Annals of Operations Research **246**(1-2), 181–203 (nov 2016)
9. Hu, S.R., Liu, C.T.: An optimal location model for a bicycle sharing program with truck dispatching consideration. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 1775–1780. IEEE (oct 2014)
10. Kloimüllner, C., Raidl, G.R.: Hierarchical Clustering and Multilevel Refinement for the Bike-Sharing Station Planning Problem. pp. 150–165. Springer, Cham (2017)
11. Lin, J.R., Yang, T.H., Chang, Y.C.: A hub location inventory model for bicycle sharing system design: Formulation and solution. Computers & Industrial Engineering **65**(1), 77–86 (may 2013)
12. Liu, J., Li, Q., Qu, M., Chen, W., Yang, J., Xiong, H., Zhong, H., Fu, Y.: Station Site Optimization in Bike Sharing Systems. In: 2015 IEEE International Conference on Data Mining. pp. 883–888. IEEE (nov 2015)
13. López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L.P., Birattari, M., Stützle, T.: The irace package: Iterated racing for automatic algorithm configuration. Operations Research Perspectives **3**, 43–58 (2016)
14. Megiddot, N., Supowits, K.J.: On the Complexity of Some Common Geometric Location Problems. SIAM Journal on Computing **13**(1) (1984)
15. Mladenović, N., Hansen, P.: Variable neighborhood search. Computers & Operations Research **24**(11), 1097–1100 (nov 1997)
16. Mladenović, N., Brimberg, J., Hansen, P., Moreno-Pérez, J.A.: The p-median problem: A survey of metaheuristic approaches. European Journal of Operational Research **179**(3), 927–939 (jun 2007)
17. Park, C., Sohn, S.Y.: An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of Seoul. Transportation Research Part A: Policy and Practice **105**, 154–166 (nov 2017)
18. Reese, J.: Methods for Solving the p-Median Problem: An Annotated Bibliography (2006)
19. Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O 'mahony, E., Shmoys, D.B., Woodard, D.B.: Predicting Bike Usage for New York City's Bike Sharing System (2015)
20. Whitaker, R.A.: A Fast Algorithm For The Greedy Interchange For Large-Scale Clustering And Median Location Problems. INFOR: Information Systems and Operational Research **21**(2), 95–108 (1983)