

K-MEDOID PETAL-SHAPED CLUSTERING FOR THE CAPACITATED VEHICLE ROUTING PROBLEM

J.H. Bührmann^{1*} & F. Bruwer¹

ARTICLE INFO

Article details

Presented at the 32nd annual conference of the Southern African Institute for Industrial Engineering (SAIIE), held from 4-6 October 2021 in Muldersdrift, South Africa.

Available online 29 Nov 2021

Contact details

* Corresponding author
joke.buhrmann@wits.ac.za

Author affiliations

¹ School of Mechanical, Industrial and Aeronautical Engineering, University of Witwatersrand, South Africa

ORCID® identifiers

J.H. Bührmann
<https://orcid.org/0000-0003-0657-9933>

F. Bruwer
<https://orcid.org/0000-0003-2811-100>

DOI

<http://dx.doi.org/10.7166/32-3-2610>

ABSTRACT

In this research, k-medoid clustering is modelled and evaluated for the capacitated vehicle routing problem (CVRP). The k-medoid clustering method creates petal-shaped clusters, which could be an effective method to create routes in the CVRP. To determine routes from the clusters, an existing metaheuristic – the ruin and recreate (R&R) method – is applied to each generated cluster. The results are benchmarked to those of a well-known clustering method, k-means clustering. The performance of the methods is measured in terms of travel cost and distance travelled, which are well-known metrics for the CVRP. The results show that k-medoid clustering method outperforms the benchmark method for most instances of the test datasets, although the CVRP without any predefined clusters still provides solutions that are closer to optimal. Clustering remains a reliable distribution management tool and reduces the processing requirements of large-scale CVRPs.

OPSOMMING

Hierdie werk ondersoek die gebruik van die k-medoid groeperings metode in die begrensde In-staat-gestelde Voertuig Roetebeplanningsprobleem (IVRP). Die k-medoid groeperingsmetode is geneig om blomblaarvormige groeperings te vorm, wat 'n effektiewe metode vir roetebeplanning in die IVRP kan voorhou. Die gekose groeperingsmetode word opgevolg deur 'n afbreek-en-herkonstrueer metode wat roetes vir elke groepering apart konstrueer. Die resultate van hierdie algoritme is vergelyk met resultate verkry deur 'n soortgelyke toepassing van die alombekende k-gemiddelde groeperings metode vir drie vooraf geïdentifiseerde datastelle. Twee maatstawwe is gebruik om die kwaliteit van die oplossings met mekaar te vergelyk, naamlik reiskostes en totale afstand afgelê. Die resultate dui daarop dat die k-medoid groeperingsmetode oor die algemeen beter resultate lewer as die k-gemiddelde groepering. Verder dui die resultate daarop dat die algoritme sonder enige vooraf opgestelde groeperings die beste resultate oplewer. Ten spyte hiervan, kan groeperingsmetodes nog steeds beskou word as 'n handige manier om grootskaalse probleme van die IVRP op te breek in kleiner subprobleme wat dan kan lei tot korter oplossingstye. Groeperingsmetodes is ook waardevol vir besluitnemers in die segmentering van groot verspreidingsnetwerke.

1 INTRODUCTION

The vehicle routing problem (VRP) is a widely studied optimisation problem in the field of physical distribution and logistics [1, 2, 3]. The objective is to find optimal routing schedules to minimise the distribution costs for one or more vehicles from one or several depots to a number of geographically scattered customers. In the constrained variant, the capacitated vehicle routing problem (CVRP), is subject to a vehicle capacity constraint [1]. The problem is considered NP hard [3], and many variants of the CVRP exist, resulting in a vast amount of literature on this class of problems in the past 50 years [1, 4].

The purpose of this study was to investigate the use of a cluster-first route-second method. The k-medoid clustering method was used to create petal-shaped clusters for the route sequencing. In order to benchmark the results of the k-medoid method, a well-known clustering method, the k-means, was selected. The

methods were applied to three datasets of different sizes to evaluate the feasibility of and ideal conditions for their application.

To address the routing component of the study, the ruin and recreate (R&R) metaheuristic method was used [5]. The method is available through an open-source software package, ODL Studio [6]. This tool was used to execute and measure the CVRP results and allowed for the geographical mapping of found solutions [7].

2 LITERATURE REVIEW

2.1 Capacitated vehicle routing problem (CVRP)

In the field of combinatorial optimisation, the CVRP remains one of the most challenging problems, in which new variants in the problem statement are continually being added to the academic literature [8]. Laporte *et al.* [3] provide a comprehensive list of the advances in solving methods since the 1950s, as well as a problem formulation, called the three-index vehicle flow formulation [1]. Laporte *et al.* [2] and Eksioglu *et al.* [4] provide very detailed and extensive surveys, including practical applications of the CVRP and the proper algorithms to tackle them. Numerous heuristic and metaheuristic methods have been proposed in recent years to solve larger scale datasets for which exact methods fail to find optimised solutions within a reasonable amount of time [9].

2.1.1 Petal clustering

A set partition heuristic to create promising vehicle routes, known as petals, to solve the CVRP was introduced by [10]. The heuristic, referred to as the sweeping algorithm, forms vehicle clusters by rotating a ray centred at the depot, which is used to generate non-overlapping petals [2, 3]. This was later extended into a more sophisticated petal algorithm in which the petals were formed by solving a set partition by Ryan *et al.* [11] and Renaud *et al.* [12].

For the creation of routes, the famous Clarke and Wright algorithm was introduced in 1964 and, with time, more sophisticated improvement heuristics and metaheuristic methods, including local and population search algorithms, were added to the literature [2].

2.1.2 Cluster-first route-second method

A heuristic, first introduced by Fisher and Jaikumar [13], used the concept to cluster customers into vehicle groups and then to construct routes for the vehicle groups. The clustering was based on seed points that became the basis of each vehicle group, and the generalised assignment problem was used to assign customers to seeds [2]. The routing sequence was then established on the basis of least insertion – i.e., customers were added to the route where the cost would be the lowest. Over time, numerous two-phase heuristics were introduced in which one phase constructed vehicle groupings or clusters and the next phase solved the routing sequence [14].

2.1.3 The ruin and recreate (R&R) method

Several route optimisation techniques were explored to address the routing component of the study. The R&R method, introduced by Schrimpf *et al.* [5], was selected because of its exceptional published results. It starts with an existing solution that can easily be created using a basic construction heuristic method. Once an existing solution has been created, a significant portion of the customers on a route are then deconstructed (*ruin*) and rebuilt using a *recreate* algorithm. The method iterates between the ruin and recreate phases in search of better solutions.

Whereas many methods try to improve an existing solution, the principle behind the R&R method is to ruin a large fraction of the current solution because this increases the degrees of freedom to restore it to a new admissible solution that still adheres to all constraints.

Different ruin modes are used; customers are removed from current routes using a radial ruin (identifying adjacent customers in the same area or time window, for example), random ruin (random selection of customers), or sequential ruin (customers in sequence on a route) [5]. Recreation is based on the best insertion strategy in which customers are randomly added one by one, each time in the best possible way without violating constraints. Once a new solution has been constructed, a decision rule is applied to determine whether or not the new solution will be accepted. The decision is based on the application of the greedy algorithm (only allow better solutions), threshold acceptance, or a simulated annealing approach in which acceptance is based on a temperature parameter.

2.2 Clustering algorithms

In this study, the focus is on two clustering methods: the k-means and the k-medoid. Both methods are iterative partitioning methods and require a k number of seeds to be specified [15].

2.2.1 K-means clustering

The best-known iterative clustering method, k-means, has widespread applications, including unsupervised learning classification problems in neural networks [16]. The clustering process begins by randomly selecting a number of seed points equal to k clusters. Every customer is allocated to a cluster by associating the customer with their nearest mean. Cluster centroids are calculated as the means of the longitude and latitude of the customers assigned to the cluster. Customers are reassigned if another cluster centroid is now nearer to the cluster than the current cluster assignment. After the reassignments, the cluster centroids are recalculated and become the new means of the clusters. Customer assignments and the recalculation of cluster centroids are repeated until no further distance savings can be made or until convergence is reached [15].

2.2.2 K-medoid clustering

The k-medoid clustering method forms part of the iterative partitioning cluster method category. The method was first introduced by Kaufman and Rousseeuw [17] as a variation on other partitioning methods. The main difference, compared with other partitioning methods, is that it works with medoids as opposed to means. The medoid of a cluster can be defined as the most centrally located point in a cluster, or the point in the cluster where the average dissimilarity from all the other points in the cluster is at a minimum. The method is often referred to as the partitioning around medoids (PAM) clustering [18].

The cluster shapes associated with the k-medoid method seem typically to segregate the geographical region into a pie-chart-type clustering around the depot, which can be used as the basis for creating petal-shaped vehicle clusters [19].

Figure 1 illustrates the difference between the k-means and the k-medoid clustering methods. In this example, the same random cluster seeds are chosen as starting cluster seeds for both methods. In step 1, customer points are assigned to their closest seeds, represented by the red and blue clusters. In step 2, the cluster seeds need to be recalculated. Step 2a) illustrates how the clusters seeds are calculated based on the means of the longitudes and latitudes of the customer points for the k-means clustering method. Step 2b) illustrates that the k-medoid clustering method selects the most central customer point as the cluster seed. These steps are then repeated until there are no more changes in customer allocation.

This example demonstrates that the k-medoid method is less sensitive to outliers, resulting, in this case, in quicker convergence. However, the k-means method is less computationally expensive than the k-medoid method because of the number of calculations needed to determine the most central customer point for each cluster in step 2 for the k-medoid method [18]. This results in the k-means method converging to a final solution faster than the k-medoid method.

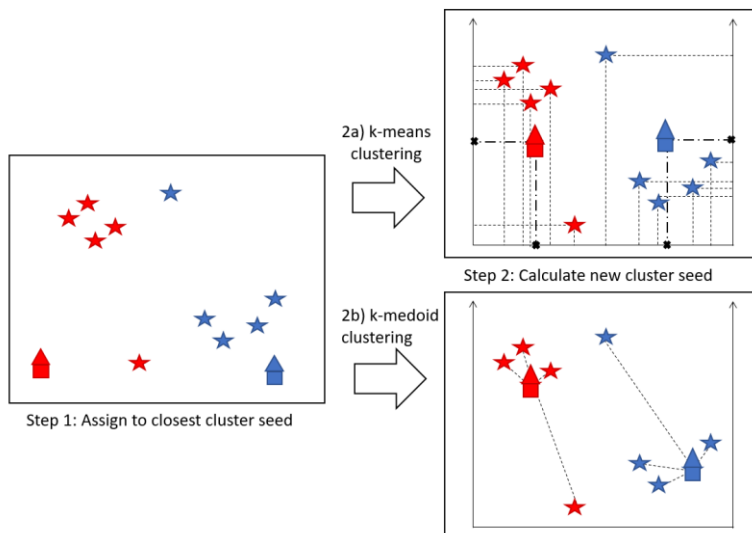


Figure 1: Example of the k-means and the k-medoid clustering methods

3 METHODOLOGY

The open-source software package ODL Studio [6, 7] was used in this study. The software makes use of the R&R method to create CVRP solutions [5]. In this software, the user can specify unique customer clustering groups either according to geographic location or delivery windows or left as individual distribution points for the R&R method to construct solutions freely.

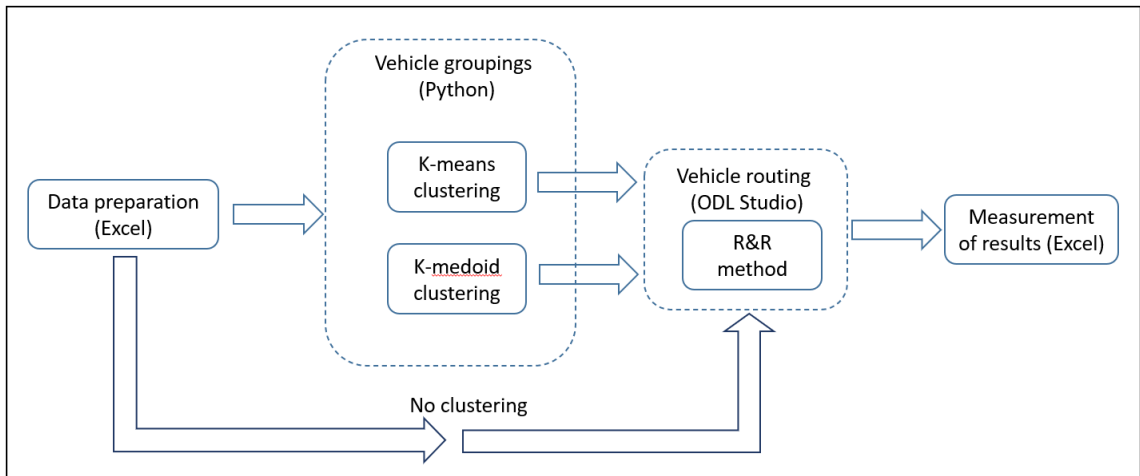


Figure 2: Research method overview

Figure 2 provides an overview of the research method used in this study. Three datasets, defined below, were tested. First, the data was prepared in Excel. Next, three scenarios were tested for each dataset. In the first two scenarios, vehicle groupings were created using either the k-means or the k-medoid clustering method coded in Python, and the R&R method was used to create the routes using ODL Studio. In the third scenario, no predefined clustering was used, and the R&R method was allowed to create its own vehicle groupings while also constructing routes. The results of the three scenarios were then compared, based on the measures defined in Section 3.2.

3.1 Building the model

The purpose of the designed method was to measure the effectiveness of the k-medoid clustering for the CVRP. Although several algorithms and software applications are available to address the routing section of the problem, clustering tools beyond k-means clustering for distribution network problems are scarce. Therefore, the clustering method had to be developed and tested by the authors. The cluster results, along with other CVRP parameters, were then provided as input to an existing routing software package that determined the routes of the vehicles to the customer locations. Thus, a cluster-first route-second approach was applied.

It was decided to use k-means clustering as a benchmark, as this is a relatively simple and widely used clustering method that is significantly covered in the literature. K-means is a popular clustering algorithm that has widespread application in data science, more recently particularly in machine learning problems. The k-means function is included in the scikitlearn module of Python [20] and could easily be imported into the model. This proved to be a strong benchmark method for this study.

3.2 Input data and data collection

Three different datasets of varying sizes and locations were used to ensure that unbiased conclusions could be reached at the analysis stage. The number of clusters was also varied per dataset to assess the effect of cluster sizes on the results.

The datasets were defined as follows:

- Dataset A: United Kingdom (size 100)
- Dataset B: Germany (size 500)
- Dataset C: Austria (size 1500)

All three datasets were sourced from ODL Studio’s demo data library [2]. Each dataset consisted of the customer locations with latitude and longitude values, the demand per customer, the vehicle capacity (=500), cost per kilometre (=0.001), and fixed cost per vehicle (=100). The customer locations with their associated cluster indices, as determined by the clustering program, were provided as input to the routing software.

The method was executed for different scenarios to compare the impact of the three different clustering algorithms on the final results. For the three datasets, the number of clusters (k) was incremented for each dataset. For every iteration, the following performance metrics were captured in Microsoft Excel:

- Total travelling costs (combination of fixed cost and variable cost per kilometre) can be derived from the objective function formula for the VRP and adding a fixed cost term for vehicles [1, 19]:

$$\sum_{j=1}^K f y_k + \sum_{j=0}^N \sum_{i=0, i \neq j}^N c_{ij} x_{ijk} \tag{1}$$

where, the first term represents the vehicle fixed costs, with: M is the total number of vehicles available, f is the fixed cost per vehicle, y_k is the binary decision variable used to determine if the vehicle is being used; and

the second term represents the VRP variable cost: N is the total number of customers; c_{ij} is the costs to travel between two points i and j with the depot being represented as point 0 and the customer points, points 1 to N ; and x_{ijk} is the binary decision variable used to represent a connection between points i and j on a route for the vehicle k .

- Total distance travelled using the formula [1]:

$$\sum_{j=0}^N \sum_{i=0}^N d_{ij} x_{ij} \tag{2}$$

where d_{ij} is the Euclidean distance between points i and j , and x_{ij} is the binary decision variable used to represent the connection route segment between points i and j on any of the routes. The rest of the variables are the same as defined in Eq. (1).

- Possible violation of constraints (e.g., unassigned customers or vehicle capacity violation)

Finally, the generated routing solution was analysed visually on the geographical map. The findings and the analysis of the results are discussed in section 4.

4 RESULTS

The three different datasets were clustered by implementing the k-medoid method and the benchmark method, k-means, over a range of k-values. After the clustering results had been imported into the routing software, route sequences were calculated, and the resulting CVRP performance metrics were recorded for every iteration.

The k-medoid method showed promising results compared with the benchmark method, k-means. In most instances, better results were obtained for the total distance travelled and the total cost. The performance metrics and a visualisation of the generated k-medoid clusters are displayed in Tables 1–3 and Figures 3a–3c for each dataset. The typical petal-shaped clusters created by the k-medoid method are especially visible in Figures 3a and 3b.

Table 1: Clustering results for dataset A: United Kingdom (size=100)

Number of clusters	k-medoid travel costs	k-mean travel costs	k-medoid distance (km)	k-means distances (km)
3	27.43	27.91	2043	2079
4	28.85	31.37	2149	2336
5	30.52	31.59	2273	2353
6	32.64	33.66	2431	2507

Table 2: Clustering results for dataset B: Germany (size=500)

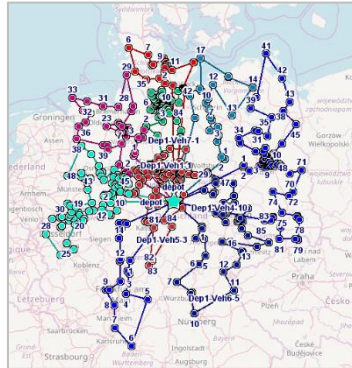
Number of clusters	k-medoid travel costs	k-mean travel costs	k-medoid distance (km)	k-means distances (km)
5	121.51	126.09	9049	9450
6	124.48	125.02	9270	9311
7	125.59	126.92	9353	9452
8	131.06	132.11	9761	9839

Table 3: Clustering results for dataset C: Austria (size=1500)

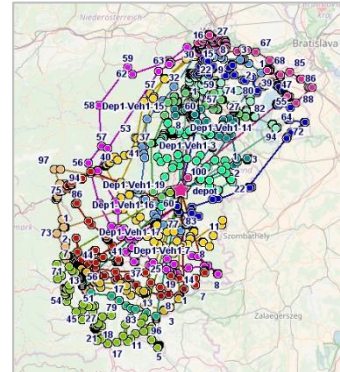
Number of clusters	k-medoid travel costs	k-mean travel costs	k-medoid distance (km)	k-means distances (km)
12	57.68	58.49	4296	4356
13	57.24	58.16	4263	4331
14	58.28	58.40	4341	4349
15	58.69	59.58	4371	4437



3a) Dataset A (UK, k=3)



3b) Dataset B (Germany, k=5)



3c) Dataset C (Austria, k=13)

Figure 3: K-medoid clustering results for the various datasets

When no predefined clusters are specified, ODL Studio [2] generates significantly better results. Adding clustering rules to the CVRP adds extra constraints to the problem that limits the R&R method and, as a result, weakens the result. The R&R results with no predefined clusters are compared with the clustering methods for which the number of clusters showed the best results. The results for the three datasets are displayed in Tables 4–6.

Table 4: Routing results for dataset A: United Kingdom (size=100)

Clustering technique	Clusters	Travel costs	Distance (km)
k-medoid	3	27.43	2043
k-means	3	27.91	2149
R&R method	1	25.05	1865

Table 5: Routing results for dataset B: Germany (size=500)

Clustering technique	Clusters	Travel costs	Distance (km)
k-medoid	5	121.51	9049
k-means	5	126.90	9450
R&R method	6	105.96	7892

Table 6: Routing results for dataset C: Austria (size=1500)

Clustering technique	Clusters	Travel costs	Distance (km)
k-medoid	13	57.24	4263
k-means	13	58.16	4331
R&R method	15	57.82	4306

5 ANALYSIS

Overall, the data showed that k-medoids outperformed the results of k-means in the CVRP. One could conclude, therefore, that k-medoid clustering could be successfully applied in instances where logically clustered solutions are preferred over pure metaheuristic methods.

5.1 Processing time

The processing times of both clustering methods were relatively short. The processing time for clustering never exceeded eight seconds, even for the larger dataset of 1500 data points. The clustering algorithm allows for scalability and is robust to larger problems in terms of runtime. The routing optimisation in ODL Studio [6] ranged between 15 and 45 seconds for datasets A and B, while for dataset C the processing time increased dramatically. An attempt to model an experimental dataset of 2000 customers could not be processed. For larger customer networks, more advanced processing capability would be required.

5.2 Number of clusters

The number of clusters strongly influenced the overall results. An important aspect to consider in the application to CVRPs is the capacity or size of the vehicle. Dondo and Cerdá [21] refer to ‘feasible clusters’, meaning that all the customers in one cluster should be served by a single vehicle. When a particular cluster is too large, an additional vehicle needs to be allocated to serve its customers. This adds an unnecessary constraint to the routing optimisation method, as the clusters are then subdivided.

Note that the cost to service the network, as well as the total distance travelled, increases with the number of pre-defined clusters. This is because additional vehicles have to be added or, in some cases, the vehicles are under-utilised. Figure 4 illustrates how the total cost, based on Eq. (1), increased with the number of clusters for dataset A.

5.3 Practical observations and limitations

Both clustering methods provided more logical groupings of customer locations than the routing metaheuristic determined without clustering inputs. This is an important aspect to bear in mind, as intuitive solutions are generally preferred by customers over abstract solutions [5].

From an operational point of view, the spatial grouping of customers is an important management tool. Clustering can aid the planning and assignment of resources to specific regions and customers. Decision-making with regards to the assign the vehicle fleet and drivers could be directed by using clustering methods. The addition of new customers to the network can easily be done in a clustered network by allocating the customers to specific regions (clusters) without re-running a routing algorithm.

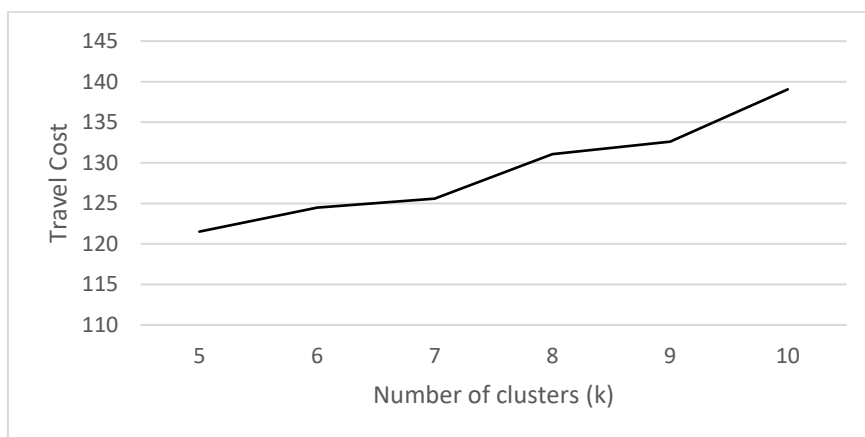


Figure 4: Total cost of CVRP compared with the number of clusters

Clustering allows the specification of a particular number of clusters. This could be valuable in distribution problems when a number of management areas or regions need to be defined.

However, for vehicle routing decisions in which the sole aim is to minimize the travelling cost and the distance travelled, neither k-medoid nor k-means clustering could outperform a modern metaheuristic method such as the R&R method, as shown by Tables 1–3. The clustering results added an extra constraint to the metaheuristic method that decreased its performance in terms of cost and distance travelled.

6 CONCLUSION

In the context of day-to-day vehicle routing decisions, the use of clustering in conjunction with modern metaheuristics seems to be beneficial, when the main objective is merely to find minimum cost solutions. Advanced metaheuristics, such as the R&R method used in this study, offered better optimisation solutions in terms of cost and distance when used without assigning predefined groupings. Adjustments could be made to metaheuristic routing methods to make them more intuitive and logical to implement.

The results suggested that the application of k-medoid, and possibly other clustering methods, should be limited to high-level distribution network planning and to assigning resources. During the modelling of the studied methods, it was clear that clustering enables effective scalability of processing times and could be advantageous for large datasets.

7 RECOMMENDATIONS FOR FUTURE RESEARCH

The exceptional results of the R&R principle presented in this work could also help to solve other optimisation problems. Specifically, it is recommended to test the results of the R&R method on more realistic variants of the VRP with additional constraints, such as the VRP with time-windows (VRPTW).

The effect of clustering on the initial design of distribution networks – e.g., determination of vehicle fleet size, geographical location of depot, and assignment of specific resources to specific regions – should be evaluated. This could also include the determination of the optimal number of clusters. Methods or techniques that are specifically relevant to distribution networks should be investigated. Given the strain of large datasets on processing resources, the impact of clustering on distribution network problems for massive datasets should be explored. Iterative clustering methods could be used to reduce the time complexity of the methods to solve these problems.

REFERENCES

- [1] G. Laporte, "The vehicle routing problem: An overview of exact and approximate algorithms," *European Journal of Operational Research*, vol. 59, no. 2, pp. 345-358, 1992.
- [2] G. Laporte, M. Gendreau, J.Y. Potvin and F. Semet, "Classical and modern heuristics for the vehicle routing problem," *International Transactions in Operational Research*, vol. 7, no. 4-5, pp. 285-300, 2000.
- [3] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, no. 4, pp. 408-416, 2009.
- [4] B. Eksioğlu, A.V. Vural and A. Reisman, "The vehicle routing problem: A taxonomic review," *Computers & Industrial Engineering*, vol. 57, no. 4, pp.1472-1483, 2009.
- [5] G. Schrimpf, J. Schneider, H. Stamm-Wilbrandt and G. Dueck, "Record breaking optimization results using the ruin and recreate principle," *Journal of Computational Physics*, vol. 152, no.2, pp. 139-171, 2000.
- [6] ODL Studio, "Open door logistics, v1.4.1," 2017 [Online] Available: <https://www.opendoorlogistics.com/downloads/> [Accessed: June 2, 2021].
- [7] P. Welch, "How to build a realtime vehicle route optimiser," White Paper, June 2017, [Online] Available: <https://www.opendoorlogistics.com/assets/pdfs/How-to-build-a-real-time-vehicle-route-optimiser.pdf> [Accessed: June 2, 2021].
- [8] A. Moolman, K. Koen and J. v.d. Westhuizen, "Activity-based costing for vehicle routing problems," *South African Journal of Industrial Engineering*, vol. 21, no. 2, pp. 161-171, 2010.
- [9] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268-308, 2003.
- [10] B.E. Gillett and L.R. Miller. "A heuristic algorithm for the vehicle dispatch problem," *Operations Research*, vol. 22, no. 2, pp. 340-349, 1974.
- [11] D.M. Ryan, C. Hjorring and F. Glover, "Extensions of the petal method for vehicle routing," *Journal of the Operational Research Society*, vol. 44, no. 3, pp. 289-296, 1993.
- [12] J. Renaud, F.F. Boctor and G. Laporte, "An improved petal heuristic for the vehicle routing problem," *Journal of the Operational Research Society*, vol. 47, no. 2, pp. 329-336, 1996.
- [13] M.L. Fisher and R. Jaikumar, "A generalized assignment heuristic for vehicle routing," *Networks*, vol. 11, pp.109-124, 1981.
- [14] J.F. Cordeau, G. Laporte, M.W. Savelsbergh and D. Vigo, "Chapter 6: Vehicle routing," in C. Barnhart and G. Laporte, eds, *Transportation, Handbooks in Operations Research and Management Science*, vol 14, Amsterdam, Elsevier North-Holland, pp. 367-428, 2007.
- [15] B. Everitt, S.L. Landau, M. Leese and D. Stahl, *Cluster analysis*, 5th ed., Chichester, United Kingdom, John Wiley, 2011.
- [16] J. Han and M. Kamber, *Data mining: Concepts and techniques*, 2nd ed., Waltham, Morgan Kaufman Publishers, 2006.
- [17] L. Kaufman and P. Rousseeuw, "Partitioning around medoids (Program PAM)," in *Finding groups in data: An introduction to clustering analysis*, New York, Wiley, pp. 68-125, 1990.

- [18] A. Jain, "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, vol. 31, pp. 651-666, 2010.
- [19] J. Bührmann, "The effects of clustering on the capacitated location-routing problem," PhD thesis, School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, 2016.
- [20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2012. [Online]. Available: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf> [Accessed: June 2, 2021].
- [21] R. Dondo and J. Cerdá, "A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows," *European Journal of Operational Research*, vol. 176, no. 3, pp. 1478-1507, 2007.