Pamukkale Univ Muh Bilim Derg, 30(1), 53-62, 2024

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

# Path inference implementing the cluster path covering method

## Küme yolu kaplama yöntemi uygulaması ile yol çıkarımı

*Kadir AKGÖL[1*], Emre DEMİR[2], İbrahim AYDOĞDU[3]*

[1]Department of Civil Engineering, Faculty of Engineering, Giresun University, Giresun, Turkey.
kadirakgol@hotmail.com.tr

[2]Department of Civil Engineering, Faculty of Engineering, Antalya Bilim University, Antalya, Turkey.
emre.demir@antalya.edu.tr

[3]Department of Civil Engineering, Faculty of Engineering, Akdeniz University, Antalya, Turkey.
aydogdu@akdeniz.edu.tr

Received/Geliş Tarihi: 25.10.2022 Revision/Düzeltme Tarihi: 08.02.2023 doi: 10.5505/pajes.2023.09522
Accepted/Kabul Tarihi: 22.02.2023 Research Article/Araştırma Makalesi

**Abstract**

*Determination of the optimal route in transportation activities is one of the major problems in transportation. Therefore, efficient techniques deserve our utmost attention to detect optimal routes. In this study, a novel method called Cluster Path Covering (CPC) has been developed and introduced to identify a route based on a sequence of location points on a network. There are already models to minimize the total path cost between the pair of nodes following a kind of sequence. However, our method aims to minimize the path cost, including the neighbourhood accessibility of the path nodes on the network. One of the major challenges for the new model is to reveal the accessibility costs between the nodes. The methodology presents the CPC method clustering the location points on a network and indicating the optimum point for each cluster. Then, the CPC method generates the best path by connecting the specific location points representing the clusters. Moreover, the shortest covering of the neighbourhood path problem (SCNPP) is introduced in this study. The novel CPC method is utilized for SCNPP, a distinctive version of the shortest covering path problem (SCPP). The performance of the CPC method is then tested on two different benchmark networks. According to the results, it provides robust and efficient outcomes for decreasing the routes' transportation costs (e.g., distances). The issues that can be solved via the CPC method include the accessibility costs of public transportation paths and the locations of stops by minimizing the costs.*

**Keywords:** Network design, Covering path, K-means, Urban planning.

**Öz**

*En uygun güzergâhın belirlenmesi ulaşım faaliyetlerindeki en büyük sorunlardan biridir. Bu nedenle, en uygun rotaları tespit etmek için etkili teknikler kullanılmalıdır. Bu çalışmada, bir ağ üzerindeki konum noktaları dizisine dayalı bir rotayı tanımlamak için Küme Yolu Kaplama (CPC) adı verilen özgün bir yöntem geliştirilmiş ve tanıtılmıştır. Bir dizi üzerindeki düğüm noktaları çiftleri arasındaki toplam yol maliyetini en aza indirecek modeller hâlihazırda bulunmaktadır. Ancak bizim yöntemimiz, ağdaki düğümlerin komşu olduğu muhite erişilebilirliğini de göz önünde bulundurarak yol maliyetini en aza indirmeyi amaçlamaktadır. Bu yeni yöntem için en büyük zorluklardan biri, düğümler arasındaki erişilebilirlik maliyetlerini ortaya çıkarmaktır. Metodolojimiz, bir ağ üzerindeki düğüm noktalarını kümeleyen ve her küme için optimum noktayı gösteren CPC yöntemini sunmaktadır. Bu yöntem, kümeleri temsil eden düğüm noktalarını birbirine bağlayarak en iyi yolu oluşturmaktadır. Ayrıca bu çalışmada, komşu muhiti kaplayan en kısa yol probleminin (SCNPP) tanımı yapılmıştır. Yeni CPC yöntemi, kaplayan en kısa yol probleminin (SCPP) ayırt edici bir versiyonu olan SCNPP için kullanılmaktadır. İlaveten, CPC yönteminin performansı iki farklı kıyaslama ağında test edilmiştir. Sonuçlara göre CPC yöntemi, güzergâhların ulaşım maliyetlerini (mesafeler gibi) azaltmak için güçlü ve etkin neticeleri bize sağlamaktadır. CPC yöntemi ile çözülebilecek sorunlar arasında, toplu taşıma güzergâhlarına mesafe açısından erişilebilirlik maliyetlerinin azaltılması ve durakların konumları arasındaki mesafelerin minimize edilmesi yer alabilir.*

**Anahtar kelimeler:** Yol ağ tasarımı, Kaplayan yol, K-means, Kentsel planlama.

## 1 Introduction

Route optimization in transportation activities is one of the major transportation problems to be cleared and offered solutions in recent years. To find out the solutions to such problems, many models and techniques are developed and introduced. There are many different solution methods according to the type of problems. For example, if the aim is only to find the shortest path between two points, a solution can be produced with methods such as Dijkstra, Bellman-Ford or Floyd Warshall algorithms. Peng, et al. [1] the emergency evacuation routes of cruise ships with the Dijkstra algorithm; Liu, et al. [2] the routes of unmanned aerial vehicles with the Bellman-Ford algorithm; Das, et al. [3] solved the routing of the semi-directed graph with the Floyd Warshall algorithm.

If the routing problem is to find the shortest path through the nodes on a network, then the problem becomes the traveling salesman problem, and the solution methods change. Inferring network paths using a dataset including spatially referenced points on a network topology is a common problem in the literature. One common problem, Traveling Salesman Problem (TSP) has been studied for a long time and has become a viable solution for clustering issues [4]-[6]. Such studies contribute invaluable algorithms to solve several path-determination problems in transportation. Moreover, studies dealing with transportation behaviour have collected spatial data to infer paths and find the most suitable routes for their research scopes and purposes. In common sense, between the origin and destination, the selection of routes by travellers or commuters is generally based on the network topology. Further, they commonly use the least cost or the shortest paths between

*Corresponding author/Yazışılan Yazar

locations on a transportation network [7],[8]. In many cases, while the distances between a pair of nodes or locations on a network are determined by the shortest distance path between them [9], typical constraints are considered to create the shortest path between the nodes in some cases [10]. Additionally, the number of arcs or nodes is limited to be included, or specific arcs and nodes are mandated to be visited by the travelling unit, and the distances between a pair of nodes are then calculated. Thus, the shortest or the least-cost paths may pass through predetermined spatial locations [4],[11]. Besides the constraints in the network to determine the direction of network paths, the objectives of the routing problem may also affect the final route connecting the origin and destination. Paths can be defined by the objective of minimizing network costs as well as maximizing access to the network [12].

It is possible to propose solutions to traveling salesman problems by methods such as mixed integer programming or Christofides' algorithm, which are known in the literature as stochastic linear programming models. Bera, et al. [13] designed the route of the unmanned aerial vehicles as a traveling salesman problem and optimized it using Christofides' algorithm. Campuzano, et al. [14] made route optimization using mixed integer programming under the name of drone assisted variable speed asymmetric traveling salesman problem. Although stochastic methods can solve small-scale problems like these examples, they are not efficient in solving large problems. Therefore, traveling salesman problems can be solved by producing faster and more efficient results with algorithms with evolutionary infrastructures such as genetic algorithm, bee colony algorithm, which are heuristic methods [15]. Moon, et al. [16] genetic algorithm for traveling salesman problems; Wong, et al. [17], on the other hand, obtained more efficient optimization results than traditional algorithms by using the bee colony algorithm.

In traveling salesman problems, there is a constraint to pass through each node in the network at least once, while in some problem types there is a constraint to pass through enough nodes to cover the network. This creates a new type of problem called the shortest covering path problem (SCPP) [12],[18]-[21]. In the solution of these problems, the nodes are grouped according to their proximity to each other, and the shortest path is sought through fewer nodes covering the network. However, in these problems, the cost of accessing the route from neighboring nodes that are on the network but not on the route is neglected. This, in case of planning public transport routes, will cause dissatisfaction for passengers and reduce interest in public transport.

The reliability of public transport services is very important to meet the passengers' expectations and travel needs and thus satisfy them [22]-[24]. One of the most important issues in passenger satisfaction depends on the accessibility rate to public transport services [25]-[28]. At this point, the location of the network stops takes place because there is a significant relationship between the location of network stops and the accessibility to public transport services [29]. According to De Oña et al. [30], most passengers (almost 80%) access network stops on foot. Also, nearly 95% of the passengers arriving at network stops use the mode of travelling on foot to reach their destination.

Many studies worked on the location of network stops according to some criteria related to their particular themes or scopes. Ceder and Wilson [31] emphasized that along with operator interests, passenger interests should also be considered. Lesley [32] tried to optimize the spacing of network stops. Van Nes and Bovy [33] revealed that stop spacing affects passenger preferences. In recent years, research in this field is still being considered, as introduced in the work of Chien and Qin [34]. They developed an optimization model to increase the accessibility of network routes by taking various cost parameters into account. They principally considered demand density where the network nodes are close to one another according to the cost (e.g., distance) between them. Alonso et al. [35] determined the stops' locations according to different congestion levels on the transportation network, considering the social cost and the modal split models. They concluded that as the demand for stops increases, the spacing between them should decrease. Further, Moura et al. [36] proposed a model analyzing the optimal stop locations according to the macro level of network social cost and the microscopic considerations of urban routes accessing the stop locations accordingly. Huang, et al. [37] developed a multimodal transit network design methodology to locate the main stations related to railway hubs. Since not only the path optimization but also the coverage of the paths is essential, Honma and Kuby [38] optimized the locations of the hydrogen refuelling stations in a network in light of coverage and accessibility.

Since the SCPPs are complex, irregular, and have a unique structure, standard optimization methods (including known meta-heuristic and stochastic methods) are not successful in solving these problems. For this reason, developing special optimization methods is necessary due to the problem structure. These particular methods are robust techniques for solving SCPP-type problems [12],[18]-[21].

In this study, unlike the shortest covering path problem, the objective function is evaluated as the sum of the cost of the main route and the access cost of all other neighboring nodes in the covering areas to the main route. Therefore, in this study, the problem was named as the shortest covering of neighborhood path problem (SCNPP) and the cluster path covering (CPC) method was developed. In this method, by minimizing the path cost, including the accessibility of the neighbourhood to the path nodes on the network, the distances between the network nodes out of the route and the stops already on the route are minimized. It is thought that the satisfaction levels of passengers will be increased with the public transportation routes to be created with this model. As a result of this, a modal shift towards public transportation will be created and this situation is very important in terms of transportation policies.

# 2 Methodology

## 2.1 Mathematical modelling of the shortest covering of neighbourhood path problem (SCNPP)

There are many methods in the literature, to provide the least-cost paths or the shortest path from location-based point observations. One of the most common techniques is to assign the least-cost route considering all the nodes at once for the same travel purpose. Current et al. [19] studied an approach for a model to determine the least-cost path between an origin and a destination on a transportation network by considering the coverage of a cost or distance standard of the nodes in the problem. This problem is named the shortest covering path problem (SCPP) and is solved by the optimization tools of linear-integer mathematical programs. In this problem, the objective is to minimize the total path cost between the pair of nodes predetermined as an origin (i.e., starting node of the trip) and destination (i.e., the terminal point of the trip). This path traverses the network, passing the maximum coverage distance of all nodes of the network [18],[19]. The cost in SCPP is named C-dist (i.e., distance travelled by vehicle) in our study. The method in this article refers to SCPP; however, our formulation considers the access of the neighbourhood to the nodes of the path. Therefore, the problem in our research can be named as the shortest covering of the neighbourhood path problem (SCNPP). According to SCNPP, the notation of our formulation is in the following. The cost in SCNPP is expressed as the summation of C-dist and P-dist (i.e., distance traveled on foot) as seen in the objective function (1).

Notation:

| | | |
|---|---|---|
| $s$ | : | The set of nodes, |
| $i, j$ | : | Indices for network nodes, |
| $m, n$ | : | Indices for nodes in the cluster, |
| $k$ | : | Index for points requiring coverage (entire set donated K), |
| $N_j$ | : | Set of nodes $i$ incident to node $j$, |
| $N_c$ | : | Set of cluster $n$ incident to node $m$, |
| $c_{ij}$ | : | Cost of traversing arc $(i,j)$, |
| $c_{mn}^s$ | : | Cost of traversing arc $(m,n)$ for $s^{\text{th}}$ cluster, |
| $o$ | : | Index denoting the origin node, |
| $d$ | : | Index denoting destination node, |
| $\delta_{jk}$ | | The minimum distance between node $j$ and point $k$ needs coverage, |
| $\Gamma^*$ | : | Coverage standard (distance or time), |
| $\Gamma_k$ | : | Set of nodes $j \mid \delta_{jk} \leq \Gamma^*$, |
| $Z$ | : | Sum of the cost of route traversal and the cost of access to the route, |
| $X_{ij}$ | : | $\begin{cases} 1, \text{ if the arc}(i,j) \text{ is on the shortest covering path} \\ 0, \text{ otherwise} \end{cases}$ |
| $X_m^s$ | : | $\begin{cases} 1, \text{ if node } m \text{ is selected in } s^{th} \text{ cluster} \\ 0, \text{ otherwis} \end{cases}$ |
| $\eta, \theta$ | : | significance coefficient $\forall \eta, \theta \in (0,1)$, |

Objective:

$$Minimize\ Z = \eta \cdot C\text{-}dist + \theta \cdot P\text{-}dist \tag{1}$$

Subject to:

$$\sum_{j \in N_o} X_{oj} = 1 \tag{2}$$

$$\sum_{j \in N_d} X_{jd} = 1 \tag{3}$$

$$\sum_{i \in N_j} X_{ij} - \sum_{i \in N_j} X_{ji} = 0 \ \ \forall j \neq o, d \tag{4}$$

$$\sum_{i \in N_j} \sum_{j \in \Gamma_k} X_{ij} \geq 1 \ \ \forall k \neq o, d \tag{5}$$

$$X_{ij} = (0,1) \tag{6}$$

$$\sum_{m \in N_c} X_m^s = 1 \tag{7}$$

$$X_m^s = (0,1) \tag{8}$$

The objective of the SCNPP (1) is to minimize the total path cost between the pair of nodes predetermined as the origin (starting node of the trip) and destination (terminal point of the trip). The coefficients of $\eta$ and $\theta$ are related to the main route's importance and the minor route's importance, respectively. The values that they can get should be between 0.00 and 1.00. For example, the value of the coefficient $\theta$ should be close to the value of 1.00 if the cost of the distance between the alternative routes in the cluster is more important. At the same time, the coefficient $\eta$ should be close to the value of 0.00. While constraints (2) require one arc to exit the origin node intermediate to the main origin and destination, constraints (3) stipulate one arc enters the destination, which is intermediate to the main origin and destination. Constraints (4) require the conservation of flow among path nodes in a dataset. While constraints (5) ensure that each node in a dataset must be covered by at least one arc in the path with the coverage standard, constraints (6) ascertain the binary/integer restrictions on the decision variables. Constraints (7) stipulate that, after clustering nodes, the total number of selected nodes from each cluster must be equal to 1. And constraints (8) require that in case of a node in a cluster is selected, the value must be 1 for that particular node, and the value of the rest of the nodes in the cluster must be 0. The formulation can be solved by the available commercial solver programs (Matlab®); however, the sub-tours may cause unexpected distance measures leading to longer distances that are not included in the trips. Not only double-checking them but also the elimination of sub-tours during the solution process may take a considerable amount of time and effort.

*C-dist* refers to the total network distance formed by the 'selected' network points. *P-dist* refers to the total network distances between 'unselected' network points and the particular 'selected' points which are the nearest. *C-dist* and *P-dist* descriptions/explanations are given as follows:

$$C\text{-}dist = \sum_{i \in N_j} \sum_{j} c_{ij} \cdot X_{ij} \tag{9}$$

$$P\text{-}dist = \sum_{m,n \in N_c} \sum_{m} \sum_{n} c_{mn}^s \cdot X_m^s \tag{10}$$

In many cases, the spatially referenced points representing a transportation activity or activities do not perfectly correspond with the road network topology [12]. Moreover, a service standard such as the distance standard of $\Gamma^*$ in the SCPP can be used for the assignment analysis which can interpret the inferring of the routes covering a distance standard in a road network.

### 2.2 K-means clustering algorithm

Clustering a group of data members attracted many researchers to simplify and regenerate their optimization algorithms. A commonly used technique called the K-means algorithm is prepared to divide M points in N dimensions into K clusters by minimizing the sum of distance squares within each cluster [39]-[41]. In the K-means clustering algorithm, the Euclidean distances are used to compute and minimize the cost of distances between the points. Our results indicate that the bisecting K-means technique is better than the standard K-means approach and as good or better than the hierarchical approaches that we tested for a variety of cluster evaluation metrics [42].

K-means clustering has a predetermined range allowing an improvement of range decision-making in pedestrian transportation. By assigning a walking distance value based on the accessibility of the bus services, the K-means clustering can work for finding a specific area or a cluster that requires attention. Thus, the number of clusters can be found automatically according to the assignment of the range cost, such as walking distance of pedestrians or potential users of bus services. The iterative process relatively consumes less time, which creates a rapid convergence rate and allows us to find the final solution rapidly. Using K-means clustering, not only the costs of travelling such as Euclidean distances can be considered but also actual spatial costs such as network distances can be taken into account. Euclidean distances can be used as input; however, actual transportation network distances may reveal the existing characteristics of the travel paths such as the interactions between the users and their surroundings.

The path distance found by the SCP model is shorter than the path we have discovered, however, the SCP model does not take into account the walking distances to the bus stops (demand points considered in the model). Since our methodology considers the walking distances to the bus stops, the coverage area indicates the accessibility to the bus stops.

### 2.3 The method of cluster path covering (CPC)

This section explains our novel methodology (i.e. the method of CPC) in detail with the help of a formulation with its notation, several figures, and a flow chart. To start with, the figures below are introduced for creating the path according to the application of the CPC method, using a 20-point network point as a sample (Figure 1).

Figure 1(a) indicates the points as the demand points that coverage needed by a path, while Figure 1(b) designates the centroids of the clusters covering all points in the 20-point dataset. As mentioned previously, briefly K-means clustering method creates a pre-determined number of clusters from a certain number of demand points. Providing this assessment, the Euclidean distances are considered in the method. However, in this study we apply the actual geographical great-circle distance by the Haversine formula, thus we revise the K-means clustering method by introducing the Haversine formula. In Figure 1(b), the points are clustered according to the distances between one another by the application of the CPC method. Since the method of CPC lets us determine the number of clusters formed, the total number of clusters in a dataset can be known. However, in our study, we would like to assign not the number of clusters in our dataset, but the maximum distance between the demand points in the dataset. To do this, the methodological steps below should be followed (Figure 2).

Firstly, divide all the dataset members (i.e., points) into 2 groups (i.e., clusters). Compute the distances between each point in each cluster. Next, find the maximum distances between the points in each cluster. Compute the maximum distance between the maximum distances in each cluster found in step 3. Compare the distance computed in step 4 with the maximum distance requested. If the distance computed at step 4 is greater than the maximum distance requested, return to step-2, and update the number of clusters by increasing 1. If the distance computed at step 4 is smaller than the maximum distance requested, stop the clustering assessment, and keep the number of clusters with their members. Thus, clustering is completed by keeping the distances between the demand points under the value of the maximum distance requested Figure 1(b). Also, the centroids of the clusters are determined and pinned.

After applying the 7-step sub-methodology above, all the demand points are divided into a couple of clusters by the distances shorter than the value of "Γ" requested. The next task is to discover which demand point will be selected to represent the coverage of the cluster particularly Figure 1(c). Using the points selected, the path will be inferred by connecting those points. This means there are 'unselected' demand points and one 'selected' point in a cluster, and the passengers should reach the 'selected' point from the 'unselected' points in the same cluster. Thereby, this transfer distance from the 'unselected' ones to the 'selected' one or accessibility cost would like to be minimized by the CPC method. The closest point in the cluster to the centroid of the cluster in terms of distance confirms the minimum accessibility cost. Thus the centroid of the cluster is represented by the closest point to the actual centroid Figure 1(c). Figure 1(d) specifies the shortest path generated between the selected points. In such path-inferring cases, there may be points whose distances are the same as the cluster's centroid. For instance, if there are two members (i.e., points) in a cluster, the cluster's centroid must be at the mid-point of those two points. In such cases, since one of the points can represent the cluster's centroid, the path can be generated by choosing the one that makes the total path length shorter. In this manner, Figure 1(e) demonstrates the path inferred by connecting the points representing the centroids of the clusters. However, since point number 13 can make the path shorter, and at the same time, can represent the centroid of its cluster, point number 14 is omitted while generating the path Figure 1(e). Thus the final version of the path is inferred by the method of CPC. The flow chart in Figure 3 explains the approach of the CPC method. Firstly, cluster and determine the centroids of the clusters by following the steps in Figure 2. Next, discover which demand point will be selected to represent the coverage of the cluster, particularly in step 3. Then, generate the shortest path in step 4. Consider if more than one point is a candidate for representing the centroid; select the point which makes the path shorter in step 5. Finally, update the path according to step-5 end finish.

## 3 Analysis and findings

### 3.1 Application 1: 20-node network sample

A 20-point network sample has been selected to process the SCPP and SCNPP. This sample can be accessed from the study of Current et al. [19], which is solved by the algorithm of SCPP. With the help of the same sample, the technique of our study has already been discussed earlier in the Methodology part. The related information about the sample can be found in the Methodology section. The 20-point network sample is processed by the method of CPC in terms of SCPP and SCNPP. Since Current et al. produced a solution according to SCPP in their study, they only calculated the C-dist value. According to the results of Current et al., the P-dist value was calculated as 1310.00 in our study (Table 1). Accordingly, the results such as C-dist, P-dist, and total distance (i.e., C-dist plus P-dist) are discovered, as depicted in Table 1, Figure 4, and Figure 5. As the maximum service distance value is 182, the method of Current et al. provides 2261.00 and 1310.00 for C-dist and P-dist, respectively. However, the CPC method calculates 2335.00 and 1257.00 for them, respectively.

When the same example is solved with the CPC method, the *C-dist* value is larger, and the *P-dist* value is smaller. Considering both cost values have the same level of significance $\eta = 0.5; \theta = 0.5$ when solving SCNPP; although almost similar costs were found (1785.50 and 1796.00), the cost was computed slightly higher in the CPC method. Here, different situations will occur according to the choice of importance coefficients. This situation has also been studied in detail.

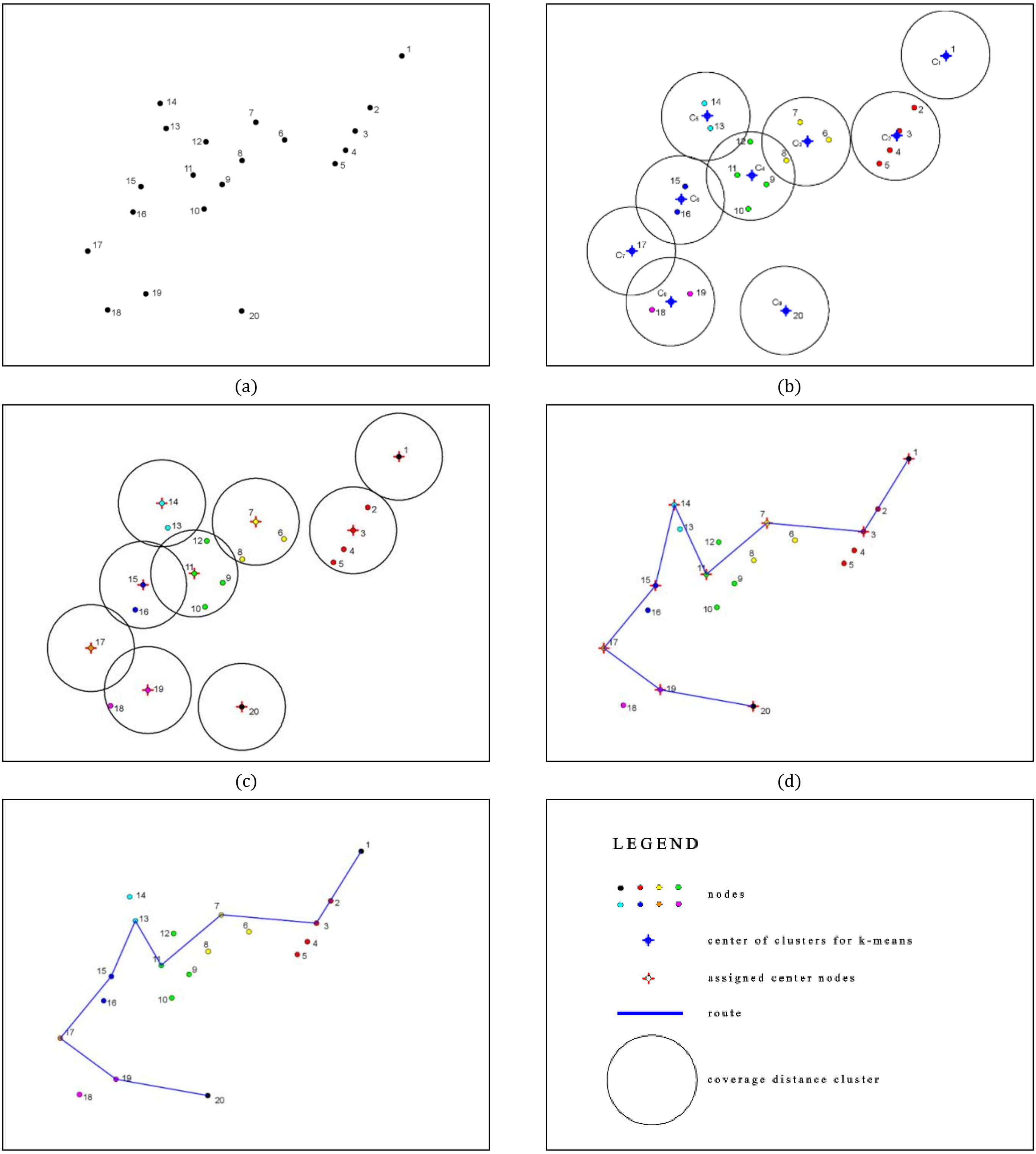


Figure 1(a): Points in the dataset. (b): Indicating the centroids of clusters. (c): Clusters and their coverages. (d): The path generation by connecting the centroid representative points. (e): The final version of the path. (f): Legend.

Table 1. Comparisons of C-dist and P-dist values for application 1.

| | | Method of Current et al. (1984) | Method of CPC |
|---|---|---|---|
| SCPP | *C-dist* | 2261.00 | 2335.00 |
| Additional | *P-dist* | 1310.00 | 1257.00 |
| SCNPP | *C-dist + P-dist* | 1785.50 | 1796.00 |

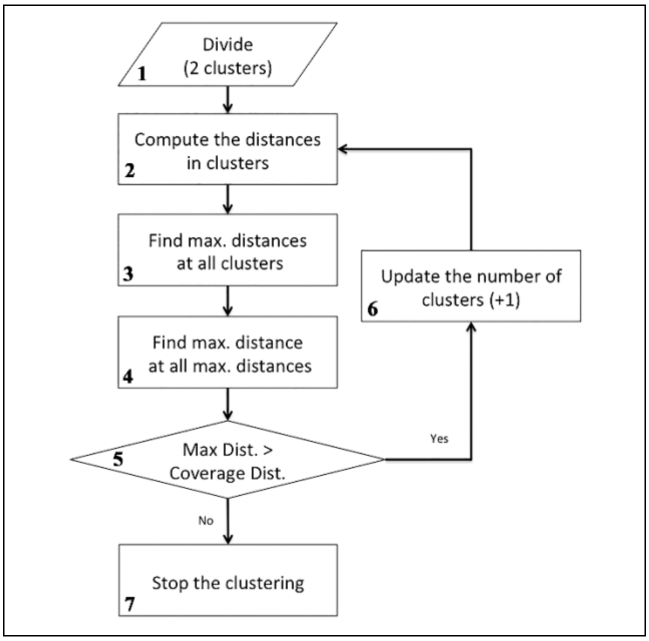


Figure 2. Clustering method flow chart.

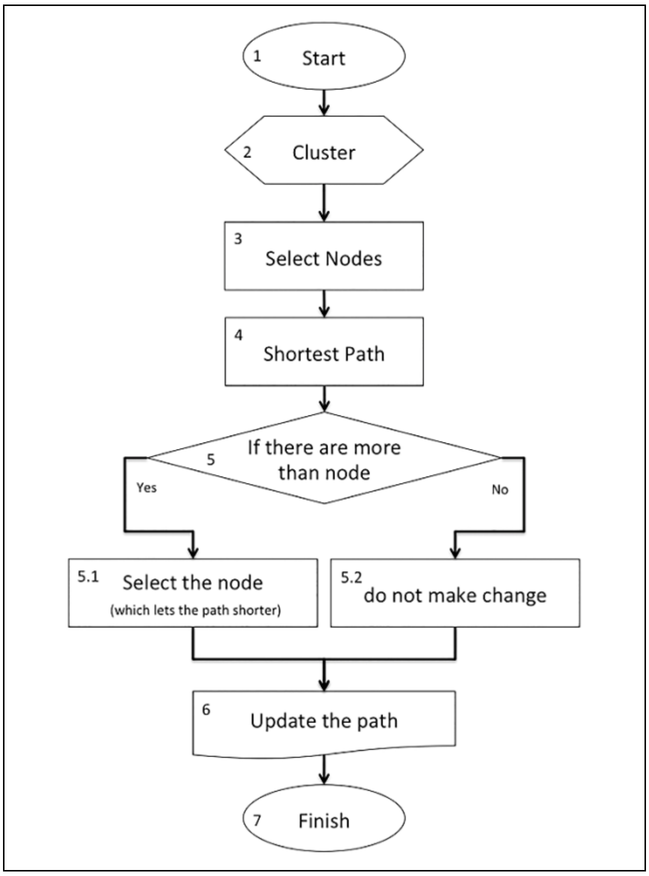


Figure 3. The approach of the CPC method.

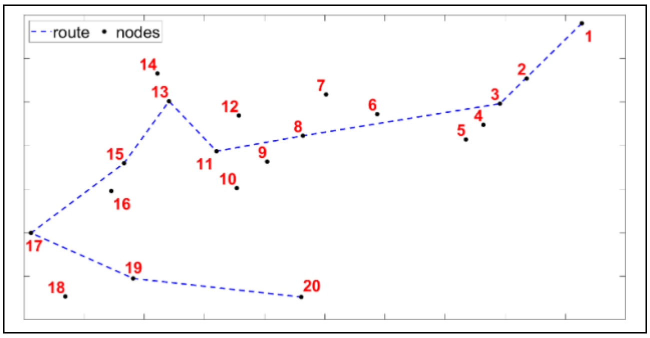


Figure 4. Best route obtained by the method of Current et al. (1984) for application 1.

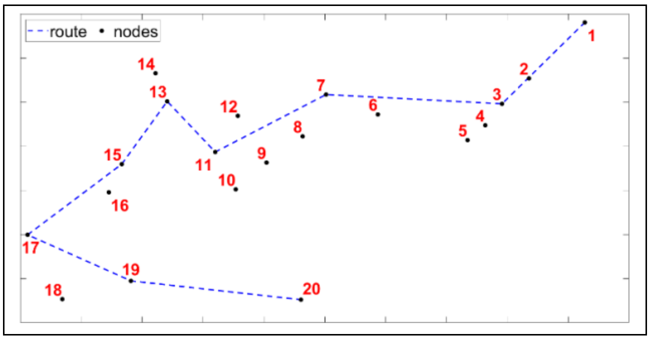


Figure 5. Best route obtained by CPC method for application 1.

### 3.2 Application 2: 55-node network sample

Additional to the first application, another application includes another sample of 55-point network nodes. In this application, the sample is taken from the study of Niblett and Church [18]. The sample network, in this case, is larger with the number of nodes, therefore the distance matrix between the nodes forming the network is greater than the sample of the first application. In their study, the coordinates of the network points were not given, therefore their data are visualized in the drawing software of Autocad®. By doing this process, the points achieved numeric values of their locations. Furthermore, the coordinates provided were calibrated to possess the solution distance of the path in the study of Niblett and Church [18]. Lastly, the travel path distances between the network points were computed by Dijkstra's method, which delivers the shortest path. Thus, the relationships between those 55 network points (i.e., costs of distances) were achieved.

The 55-point network sample is evaluated by the method of CPC. Niblett and Church [18] produced two different paths between points 21-27 on the 55-node network according to the method they developed. C-dist values were taken from the study of Niblett and Church [18], however P-dist values were calculated in our study. Again in our study, a route optimization was made by the CPC method using the same road network and origin and destination nodes that are in the study of Niblett and Church [18]. Thus, the results such as C-dist, P-dist, and total distance (i.e., C-dist plus P-dist) are computed and shown in Table 2, Figure 6, Figure 7, and Figure 8.

Table 2. Comparisons of C-dist and P-dist values for application 2.

| | Method of Niblett and Church (2016) (1st computation) | Method of Niblett and Church (2016) (2nd computation) | Method of CPC |
|---|---|---|---|
| *C-dist* | 162.69 | 155.72 | 204.57 |
| *P-dist* | 162.08 | 193.60 | 99.21 |
| Total (*C-dist* + *P-dist*) | 162.39 | 174.66 | 156.89 |

The results in Table 2 are calculated assuming that the importance coefficients of the C-dist and P-dist costs are equal η=0.5; θ=0.5. When the path costs produced by the CPC method are examined, an economical result (156.89) has been obtained compared to both solutions of Niblett and Church (162.39 and 174.66).

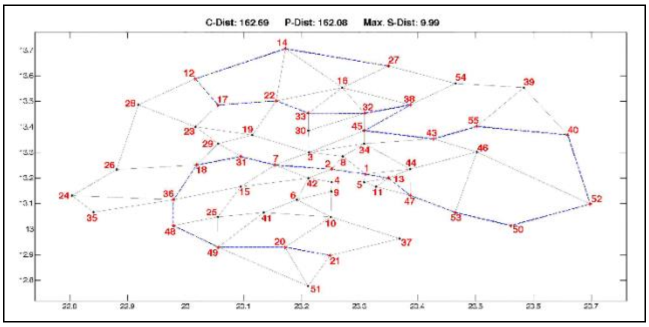


Figure 6. Best route obtained by the method of Niblett and Church [36] (1st computation) for application 2.

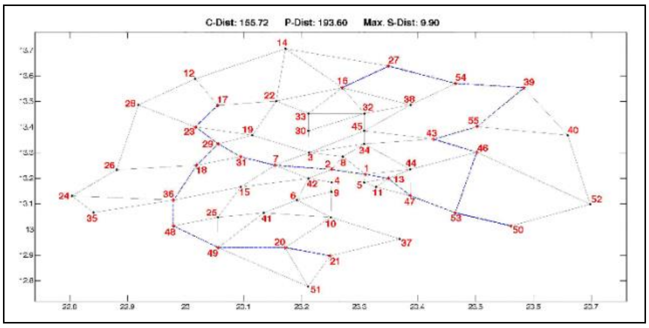


Figure 7. Best route obtained by the method of Niblett and Church [18] (2nd computation) for application 2.

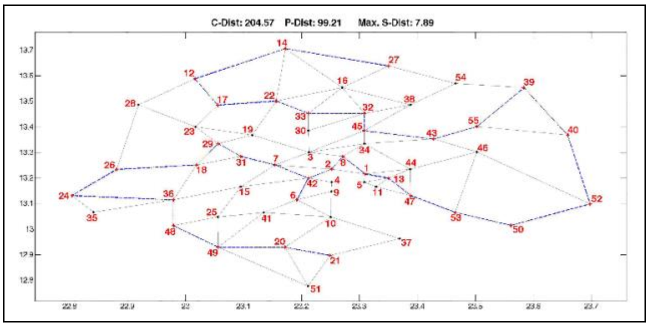


Figure 8. Best route obtained by CPC method for application 2.

### 3.3 Application 3: Testing performance of CPC

The significance coefficients added as "$\eta$" and "$\theta$" to the objective function (1) allow us to decide whether the distance travelled by vehicle or the walking distance will be minimized for the path. The paths to be inferred by taking "$\eta = 1$" and "$\theta = 0$" will minimize the total walking distance of travellers to the stops on the path. Likewise, the paths to be inferred by taking "$\eta = 0$" and "$\theta = 1$" will minimize the total distance travelled by a vehicle on the path. In addition, these coefficients can be chosen as different combinations such that they remain between "0" and "1" and the sum of the two is "1". Here, depending on the type of work, whichever distance (walking distance or distance travelled by vehicle) has higher importance level, the important factor of that particular mode of transport should be approached to 0.

In this part of the study, an application showing the results of the optimization (cost, distance, and path) based on the change in the coefficients "$\eta$" and "$\theta$" has been made. In this application, a 55-node Swain network was chosen, and optimum results were obtained using different combinations of significant coefficients on this road network. In the paths calculated for the exemplary application, node points 21 to 27 were chosen as the origin and destination points. According to the optimization results obtained, *C-dist* and *P-dist* distances are given in Figure 9. In addition, a scale that has been transitioned from the minimization of the total distance travelled by vehicle only to the minimization of the total distances covered by the walking distance has been created. Thus, it can be seen how the distances resulting from the optimization will change.

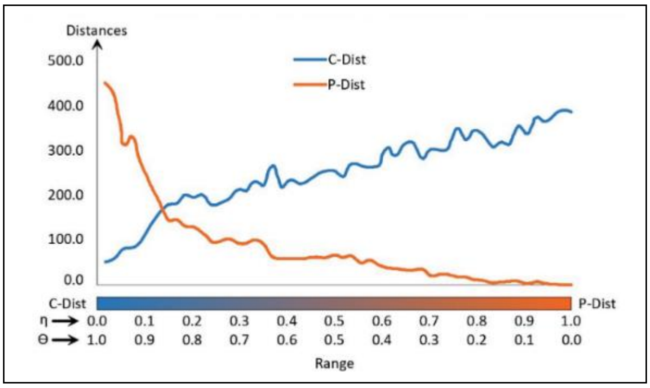


Figure 9. Distances optimized according to the change in the importance coefficients of "η" and "θ" for application 3.

Some paths inferred as a result of this optimization are given in Appendix A1. Consequently, according to the situation where the distance travelled by vehicle is in the focus and the walking distance of the pedestrians to the stops on the path is insignificant ($\eta = 0$ and $\theta = 1$), the vehicle travels 52.0936 units by choosing the shortest route in the road network. This path completed its journey by passing through 12 nodes in the road network. The sum of the shortest walking distances from the remaining 43 nodes to the points on the path is 448.3816 units. Again, the person walking the longest distance on this network will walk a distance of 26.739 units. This total network distance is basically from an 'unselected' network point to the nearest 'selected' point on the path inferred according to predetermined importance coefficients (Figure 10).

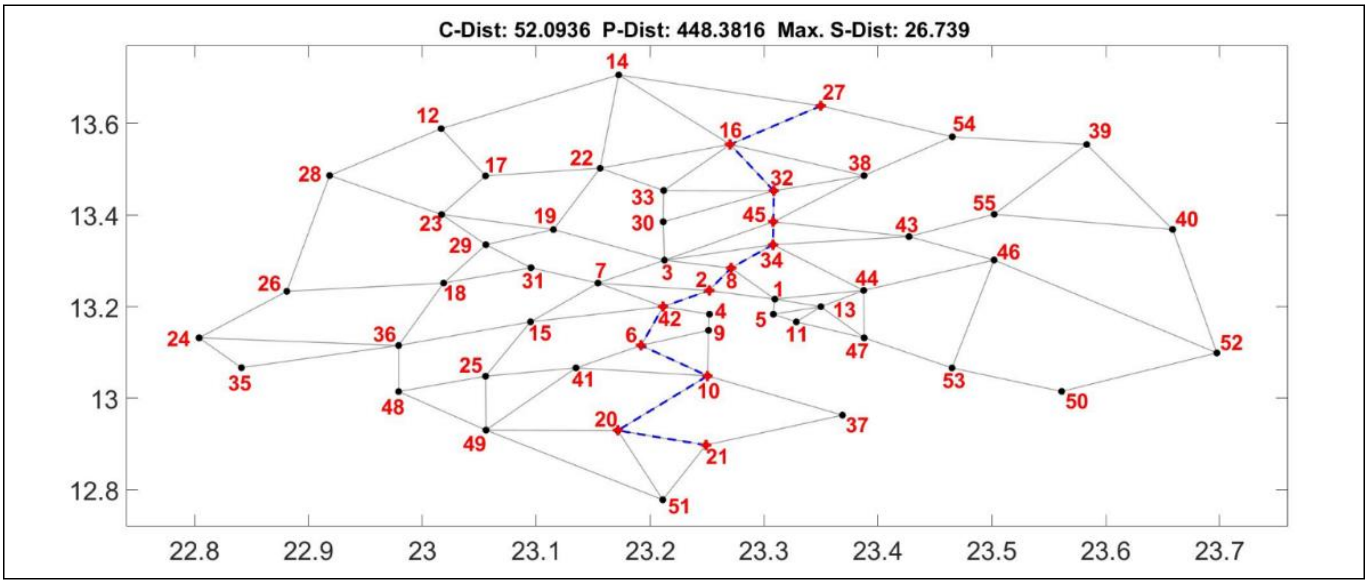


Figure 10. The path from the node 21 to 27.

According to the change of importance coefficients between 0 and 1, the change in the cost specified in the objective function is given in Figure 11. If the importance is given only to the distance travelled by the vehicle, or if the importance is given only to walking distance, the total cost becomes higher. This is a matter of preference based on how the operator tends. Of course, in this application, it is accepted that the distance travelled by vehicle and foot will cost the same. Therefore, each operator should make his cost conversion here. Considering application 3, choosing the coefficients as $\eta = 0.25$ and $\theta = 0.75$ approximately will reduce costs.

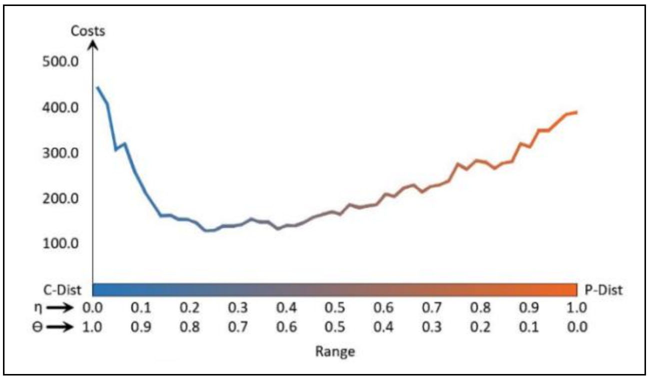


Figure 11. The change in the cost specified in the objective function (1).

## 4 Concluding remarks and future work

According to SCNP, although a longer distance is achieved by the method of CPC, the difference is only 0.6% (Table 1). For this reason, one can suppose that the method of CPC has shown a feasible performance for the case of the 20-point network. When the method of CPC was applied to a network with 55 nodes (e.g., a relatively more extensive network than the sample in Table 1), both the *P-dist* and the total distance were found to be shorter than the previous method (Table 2). Hence, as the number of nodes or network points increases, the incline in the efficiency of the CPC method is observed.

As there is a relatively small network (i.e., 20-point network), the application of SCNP (*C-dist* and *P-dist* in total) gives a distance that is only 0.6% different from the distance given by the SCPP solution (Table 1). However, when the network grows, using a 55-point network as an example in Table 2, the effect of our method named CPC can be more visible. Moreover, in both cases (Table 1 and Table 2), the distances of *P-dist* are obtained shorter using the CPC method. Thus, while the efficiency of the CPC method increases as the number of network nodes increases, it is observed that the CPC method can always find the distance of the *P-dist* shorter than the previous methods. Therefore, when the coefficient θ in formulation (1) (i.e., the weight of *P-dist*) becomes larger, the CPC method's efficiency will also become larger. Overall, the method of CPC provides feasible results to infer a network path considering the network nodes around the path. The method of CPC achieves its novelty from a new approach of the shortest covering path, called SCNP, in this article.

The cluster results may affect the final path because future work is already planned as research on the observations of various coefficients or weights with different network costs. The evaluations are expected to be published with applications of several cases. Although the K-Means method is a good technique for clustering, other clustering techniques can perform better in the CPC method. In this study, the fact that the K-means method considers Euclidean distances limits the more efficient results of the model. Again, when clustering, mandatory nodes on the main route are ignored. The selection of mandatory nodes in clustering can be added as a limitation to the model. Hence, the effects of the clustering techniques for the CPC method are considered for future work.

## 5 Author contribution statement

In the study, Kadir AKGÖL and İbrahim AYDOĞDU contributed to the creation of the idea and the design. In addition, Kadir AKGÖL contributed to the literature review and analysis of the results. Finally, Emre DEMİR and İbrahim AYDOĞDU contributed to evaluating the results obtained, providing the data used, examining the results, checking the spelling, and checking the article for content.

# 6 Ethics committee approval and conflict of interest statement

"Ethics committee approval is not required for the prepared article".

"The prepared article has no conflict of interest with any person/institution".

# 7 References


[1] Peng C, Yu S, Zhang L. "Emergency evacuation route planning of cruise ship based on intelligent optimization algorithm". *Intelligent Networked Things: 5th China Conference*, *CINT 2022*, Urumqi, China, 7-8 August 2022.

[2] Liu B, Ni W, Liu RP, Guo YJ, Zhu H. "Optimal routing of unmanned aerial vehicle for joint goods delivery and in-situ sensing". *IEEE Transactions on Intelligent Transportation Systems*, 24(3), 3594-3599, 2022.

[3] Das R, Sahoo L, Samanta S, Simic V, Senapati TJM. "Identifying the shortest path of a semidirected graph and its application". *Mathematics*, 10(24), 1-13, 2022.

[4] Dantzig G, Fulkerson R, Johnson S. "Solution of a large-scale traveling-salesman problem". *Journal of the Operations Research Society of America*, 2(4), 393-410, 1954.

[5] Lin S, Kernighan BW. "An effective heuristic algorithm for the traveling-salesman problem". *Operations Research*, 21(2), 498-516, 1973.

[6] Xu R, Wunsch D. "Survey of clustering algorithms". *IEEE Transactions on Neural Networks*, 16(3), 645-678, 2005.

[7] Dijkstra EW. "A note on two problems in connexion with graphs". *Numerische Mathematik*, 1(1), 269-271, 1959.

[8] Ford JLR. *Network Flow Theory*. 1st ed. California, USA, Rand Corporation, 1956.

[9] Ortúzar JdD, Willumsen LG. *Modelling Transport*. 4th ed. New York, USA, John Wiley and Sons, 2011.

[10] Santos L, Coutinho-Rodrigues J, Current JR. "An improved solution algorithm for the constrained shortest path problem". *Transportation Research Part B: Methodological*, 41(7), 756-771, 2007.

[11] Lombard K, Church R. "The gateway shortest path problem: generating alternative routes for a corridor location problem". *Geographical Systems*, 1(1), 25-45, 1993.

[12] Demir E. Identification and Characterization of Network Paths. PhD Thesis, University of Missouri, Columbia, USA, 2013.

[13] Bera A, Misra S, Chatterjee C, Mao S. "CEDAN: Cost-effective data aggregation for UAV-Enabled IoT networks". *IEEE Transactions on Mobile Computing*, 22(9), 5053-5063, 2022.

[14] Campuzano G, Lalla-Ruiz E, Mes M. "The drone-assisted variable speed asymmetric traveling salesman problem". *Computers & Industrial Engineering*, 176, 1-20, 2023.

[15] Ziyadullaev D, Muhamediyeva D, Ziyaeva S, Xoliyorov U, Kayumov K, Ismailov O. "Development of a traditional transport system based on the bee colony algorithm". *E3S Web of Conferences 365*, Strazburg, France, 30 January 2023.

[16] Moon C, Kim J, Choi G, Seo Y. "An efficient genetic algorithm for the traveling salesman problem with precedence constraints". *European Journal of Operational Research*, 140(3), 606-617, 2002.

[17] Wong L-P, Low MYH, Chong CS. "A bee colony optimization algorithm for traveling salesman problem". *2008 Second Asia International Conference on Modelling & Simulation (AMS)*, Kuala Lumpur, Malaysia, 13-15 May 2008.

[18] Niblett TJ, Church RL. "The shortest covering path problem: a new perspective and model". *International Regional Science Review*, 39(1), 131-151, 2016.

[19] Current J, ReVelle C, Cohon J. "Symposium on location problems: in memory of leon cooper: the shortest covering path problem: an application of locational constraints to network design". *Journal of Regional Science*, 24(2), 161-183, 1984.

[20] Current JR, Velle CR, Cohon J. "The maximum covering/shortest path problem: A multiobjective network design and routing formulation". *European Journal of Operational Research*, 21(2), 189-199, 1985.

[21] Current JR, Schilling DA. "The covering salesman problem". *Transportation Science*, 23(3), 208-213, 1989.

[22] Cheranchery MF, Maitra B. "Priority areas of intervention for improving urban bus services: Experience in Kolkata, India". *Transportation Research Record: Journal of the Transportation Research Board*, 2634, 17-27, 2017.

[23] Das S, Pandit D. "Importance of user perception in evaluating level of service for bus transit for a developing country like India: a review". *Transport Reviews*, 33(4), 402-420, 2013.

[24] Hu KC. "Evaluating city bus service based on zone of tolerance of expectation and normalized importance". *Transport Reviews*, 30(2), 195-217, 2010.

[25] Dell'Olio L, Ibeas A, Cecín P. "Modelling user perception of bus transit quality". *Transport Policy*, 17(6), 388-397, 2010.

[26] Wall G, McDonald M. "Improving bus service quality and information in Winchester". *Transport Policy*, 14(2), 165-179, 2007.

[27] Yannis T, Georgia A. "A complete methodology for the quality control of passenger services in the public transport business". *European Transport/Trasporti Europei*, 38, 1-16, 2008.

[28] Gökaşar I, Buran B, Dündar S. "Modelling the quality of bus services by using factor analysis on urban bus satisfaction survey data: Case of IETT". *Pamukkale University Journal of Engineering Sciences*, 24(6), 1079-1086, 2018.

[29] Ruiz M, Segui-Pons JM, Mateu-LLadó J. "Improving bus service levels and social equity through bus frequency modelling". *Journal of Transport Geography*, 58, 220-233, 2017.

[30] De Oña J, De Oña R, Eboli L, Mazzulla G. "Perceived service quality in bus transit service: a structural equation approach". *Transport Policy*, 29, 219-226, 2013.

[31] Ceder A, Wilson NH. "Bus network design". *Transportation Research Part B: Methodological*, 20(4), 331-344, 1986.

[32] Lesley LJS. "Optimum bus-stop spacing". *Traffic Engineering & Control*, 17(10), 399-401, 1976.

[33] Van Nes R, Bovy P. "Importance of objectives in urban transit-network design". *Transportation Research Record: Journal of the Transportation Research Board*, 1735, 25-34, 2000.

[34] Chien SI, Qin Z. "Optimization of bus stop locations for improving transit accessibility". *Transportation Planning and Technology*, 27(3), 211-227, 2004.

[35] Alonso B, Moura JL, dell'Olio L, Ibeas Á. "Bus stop location under different levels of network congestion and elastic demand". *Transport*, 26(2), 141-148, 2011.

[36] Moura JL, Alonso B, Ibeas Á, Ruisánchez FJ. "A two-stage urban bus stop location model". *Networks and Spatial Economics*, 12(3), 403-420, 2012.

[37] Huang D, Liu Z, Fu X, Blythe PT. "Multimodal transit network design in a hub-and-spoke network framework". *Transportmetrica A: Transport Science*, 14(8), 706-735, 2018.

[38] Honma Y, Kuby M. "Node-based vs. path-based location models for urban hydrogen refueling stations: Comparing convenience and coverage abilities". *International Journal of Hydrogen Energy*, 44(29), 15246-15261, 2019.

[39] Hartigan JA, Wong MA. "Algorithm AS 136: A k-means clustering algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100-108, 1979.

[40] Hartigan JA. *Clustering Algorithms*. 1st ed. New Jersey, USA, John Wiley & Sons, 1975.

[41] Telgarsky M, Vattani A. "Hartigan's method: k-means clustering without voronoi". *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, 13-15 May 2010.

[42] Karypis MSG, Kumar V, Steinbach M. "A comparison of document clustering techniques". *KDD-2000 Workshop on Text Mining*, Boston, MA, USA, 20 August 2000.

## 8 Appendix A

Appendix A1. Some optimal routes calculated for different significance coefficient values with the CPC method.