Role of Clustering Method in Items Delivery Optimization

Mokhammad Ridwan Yudhanegara¹,*, Sapto Wahyu Indratno¹, and Kurnia Novita Sari¹

¹Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia

* mridwan.yudhanegara@s.itb.ac.id

Abstract. In the process of items delivery, the map of the destination locations is represented as a network. The network is used to facilitate vehicles for items delivery. This paper provides an overview of the importance of the clustering method in items delivery optimization. The simulation compares the minimum distance between items delivery using the clustering method and the classic method. The results show that optimization using the clustering method has the total minimum distance.

Keywords: Minimum distance, network, spectral bisection

1. Introduction

In the items delivery industry, people demand timely delivery, fast service, and competitive delivery prices. Therefore, it takes the superiority of the management of the item delivery company to manage the business competitively and systematically. It is necessary to plan the items delivery so that the costs incurred can be optimized. Therefore, optimization methods and innovations are essential for the companies involved.

What are the benefits of using optimization methods in making decisions intuitively? Engineers work to improve the initial design of strategies to enhance the operation of installed equipment to achieve the most excellent productivity, most significant profit, minimum cost, most minor energy use, etc. In the end, these points have an impact on the financial value. Financial value provides a convenient measure of the purpose for which it is set. Several techniques are needed to help in making decisions, one of which is the optimization method as a transportation modelling.

Transportation modelling looks for the cheapest way to items delivery from multiple resources to multiple destinations. The original referred to here can be a factory, warehouse or other nodes from which the items are delivered. In comparison, the intended destinations are the location or nodes of recipients of items. Multiple delivery centre locations that must meet various demand locations will incur varying costs for each distribution at different locations. This situation has contributed to the rise of delivery costs every year. Optimization is one method that can be used. The transportation method aims to determine the amount that must be sent from each source or to each destination so that the total transportation costs are minimized [1].

The optimization method is necessary to have other approaches that must be combined. Research by Yudhanegara et al. [2] provides input on cluster analysis to divide the network into several zones. So that item delivery is more effective, for example, avoiding shipments to locations that are too far away.
Cluster analysis, referred to as clustering in the future, aims to classify objects based on the characteristics of these objects [3]. Determining the number of clusters of data is a significant problem [4 - 5]. Through this clustering, we can identify the characteristics of each group or cluster [6].

The type of clustering method also needs to be considered so that the desired cluster results are under the characteristics of each cluster member. According to Yudhanegara et al. [2], the method suitable for clustering of items delivery networks is spectral bisection. In some instances, ineffective clusters were also found, such as unbalanced cluster members, for example, too few members of the one cluster while the other clusters were too large [7].

The optimization method (transportation modelling) in finding the route with the shortest distance uses modifying the vehicle routing problem. The vehicle routing problem is a development of the travelling salesman problem [8]. The optimization method (transportation modelling) in finding the route with the shortest distance uses modifying the vehicle routing problem. This method has a scope of issues in which there are problems with the number of routes for several vehicles in one or more depots that must be determined in number to spread geographically to serve all customers.

In this case, network clustering in item delivery optimization aims to find total minimum distances. The background of the research is presented in this section (Introduction). The second section (Methods) describes network clustering, optimization, and statistical tests. The third section (Results and Discussion) presents simulations for optimization using clustering and classic methods. The conclusions of the research results are shown in the fourth section (Conclusion).

2. Methods

2.1. Network Clustering

A network can generally be grouped or partitioned into subnets consisting of several similar characteristics [9]. The grouping or partitioning in a network is known as clustering [10]. One of the methods used for clustering is the spectral bisection method. Spectral bisection is a cluster search method by dividing the network into two clusters based on the eigenvector with the second smallest eigenvalue of the Laplacian matrix [11 - 13]. See Lestari et al. [14] also for finding the second smallest eigenvalue.

This method belongs to the approach without geometric information, which is not carried out in other bisection methods which generally use geometric information. In this study, network clustering was performed using a recursive spectral bisection. Through this method, clustering is carried out in stages. The first step is to cluster a network into two clusters. The next step with an iterative process is clustering from each cluster, so that 2^k, k \in \mathbb{N} clusters are obtained from a network. The clustering process used is the recursive spectral bisection algorithm [2], as shown in algorithm 1.

**Algorithm 1. Recursive of spectral bisection**

1) Determine the Laplacian matrix \( L_G \) of the network \( G \).
2) Find the Fiedler vector \( \phi_2 \) from the second smallest eigenvalue \( \lambda_2 \).
3) Calculate the median \( m_{\phi_2} \) of all components \( \phi_2 \).
4) Select \( V_1 = \{ v_i \in V : \phi_2 < m_{\phi_2} \} \) and \( V_2 = \{ v_i \in V : \phi_2 > m_{\phi_2} \} \) and if some components are equal to \( m_{\phi_2} \), then distribute the appropriate nodes to be balanced.
5) Determine the Laplacian matrix \( L_G \) of subnet \( G \).
6) Determine cluster members according to stages 2, 3, and 4.

We get the Laplacian matrix \( L_G \) of the network \( G = (V, E) \) of size \( n \times n \), from

\[
L_G = D_G - M_G, \tag{1}
\]

where \( M_G \) is adjacency matrix \( G \) and \( D_G \) is the diagonal matrix of the network \( G \) [12]. The eigenvector \( \phi_1 \) is obtained from equation (2),
\[
L_G \varphi_i = \lambda_i \varphi_i, i = 1,2, ..., n. \tag{2}
\]

As for \( \lambda_i \) is the eigenvalue obtained from equation (3), i.e.,

\[
det(M_G - \lambda I) = 0, \tag{3}
\]

where \( I \) is an identity matrix of size \( n \times n \).

2.2. Optimization

The objective function for optimizing the route with the shortest distance in this research problem is to analyse the structure of the item delivery problem. Referring to Dror and Trudeau [15], Reinelt [16], Bernardino and Paias [17] that the flow problems with the shortest distance used in this study are:

a) One vehicle unit delivers items to several locations (nodes), but each node is only visited once.
b) The vehicle must return to the starting node (warehouse).
c) The goal is to find the route with the shortest distance.

The variables used in the objective function are

\[
\eta_{i,j}^{(q)} = \begin{cases} 
1, & \text{if vehicle } q \text{ moves to node } v_j \text{ directly from node } v_i \\
0, & \text{elsewhere,}
\end{cases}
\]

\[d_{i,j} = \text{length of delivery distance from the node } v_i \text{ to node } v_j.\]

The goal is to minimize the function

\[
Z = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{q=1}^{h} d_{i,j} \eta_{i,j}^{(q)}, \tag{4}
\]

subject to

\[
\sum_{i=0}^{n} \sum_{q=1}^{h} \eta_{i,j}^{(q)} = 1, j = 1,2, ..., n, i \neq j, \tag{5}
\]

\[
\sum_{j=0}^{n} \sum_{q=1}^{h} \eta_{j,i}^{(q)} = 1, i = 1,2, ..., n, j \neq i. \tag{6}
\]

Equation (5) guarantees that the vehicle that will deliver items to one node originates from the previous node. Then equation (6) ensures that the vehicle that has delivered the items from one node, then the vehicle must deliver the items to the next node. Equation (5) and equation (6) also indicate that each node can only be made exactly one stop by one vehicle. Then the other constraints used for equation (1) are

\[
\sum_{j=1}^{n} \eta_{0,j}^{(q)} = 1, q = 1,2, ..., h, \tag{7}
\]

and

\[
\sum_{j=1}^{n} \eta_{j,0}^{(q)} = 1, q = 1,2, ..., h, \tag{8}
\]

where 0 is the warehouse. Equation (7) indicates that each vehicle travel route starts from the warehouse, while and equation (8) suggests that every vehicle travel route end at the warehouse.
Furthermore, the nearest-neighbour heuristic method is used to solve the delivery distance optimization problem [18]. The algorithm to solve the nearest-neighbour heuristic method is described in algorithm 2.

**Algorithm 2. Nearest-neighbour heuristic**

1) Determine the starting node (warehouse).
2) Create a distance matrix.
3) The process of moving nodes is done by looking between nodes with the shortest distance. Whenever it reaches a node, it will choose the node that has not been visited with the shortest distance.
4) The optimal calculation adds up the transmission distance between nodes from the starting node through all nodes and back to the starting node.

2.3. Statistics Test

Let $D$ be a random variable that the stating distance between two nodes. For testing the difference between two means of total minimum distances, between two paired samples, it is assumed that $H_0: \mu_1 - \mu_2 = 0$ and $H_1: \mu_1 - \mu_2 < 0$, where $\mu_1 \approx \bar{d}_1$ for clustering method and, $\mu_2 \approx \bar{d}_2$ for classic methods. So, the statistical test used is the $t$-test [18]. The $t$-test formula is

$$t_{stat} = \frac{\bar{d}}{Sd \sqrt{n}}$$

(9)

where

$$\bar{d} = \frac{\sum d_i}{n},$$

(10)

and

$$Sd = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n - 1}},$$

(11)

d_i = D_1 - D_2. The decision-making criteria are $H_0$ is accepted if $t_{stat} \geq -t_{\alpha/2} \sqrt{n-1}$, and $H_0$ is rejected if $t_{stat} < -t_{\alpha/2} \sqrt{n-1}$.

3. Results and Discussion

Define the graph is the items delivery network. It is a map of destination locations. The simulation carried out is recursively using Python software to cluster the items delivery network with the spectral bisection method. The adjacency matrix of the network is presented in figure 1.
Next, the items delivery network from the adjacency matrix with edge weights is presented in figure 2. Suppose the point with label 0 is designated as warehouse, and another number is the delivery destination location. Then the network is clustered into two parts, as in the first simulation in figure 2.

**Figure 1.** Adjacency matrix for the network in simulation

This simulation presents 50 networks with an adjacency matrix in figure 1 with different edge weights. The structure of the edge weights on the first network, the second network, up to the fifty networks is other. The following steps are

a) Calculate the minimum distance travelled by vehicle 1, vehicle 2, and the minimum total distance of the two vehicles based on equation (4). Each vehicle's route is obtained by first clustering the network into two partitions and then using algorithm 2 to find the minimum distance. This step is referred to as the clustering method.

b) Calculate the minimum distance travelled by vehicle 1, vehicle 2, and the total minimum distance of the two vehicles based on equation (4). Then, the route of each vehicle is obtained by applying algorithm 2 to find the minimum distance. This step is referred to as the clustering method.

c) Compare the minimum distance obtained from each event.
The results of steps a and b are presented in table 1. The plots of minimum distance can be seen in figure 3.

Table 1. Minimum distance (in meters) from simulation

| Simulation | Clustering Method | Classic Method | Clustering method | Classic method |
|------------|-------------------|----------------|-------------------|----------------|
|            | Vehicle 1 | Vehicle 2 | Total | Vehicle 1 | Vehicle 2 | Total | Vehicle 1 | Vehicle 2 | Total |
| Sim 1      | 43       | 38       | 81    | 50       | 96       | 146   | 46       | 38       | 84    |
| Sim 2      | 56       | 43       | 99    | 54       | 45       | 99    | 43       | 38       | 81    |
| Sim 3      | 67       | 48       | 115   | 68       | 52       | 120   | 44       | 44       | 88    |
| Sim 4      | 46       | 41       | 87    | 46       | 42       | 88    | 43       | 38       | 81    |
| Sim 5      | 35       | 44       | 79    | 44       | 45       | 89    | 46       | 52       | 98    |
| Sim 6      | 44       | 39       | 83    | 52       | 86       | 138   | 31       | 38       | 69    |
| Sim 7      | 51       | 42       | 93    | 53       | 96       | 149   | 33       | 57       | 90    |
| Sim 8      | 43       | 38       | 81    | 50       | 36       | 86    | 54       | 52       | 106   |
| Sim 9      | 39       | 38       | 77    | 40       | 39       | 79    | 70       | 44       | 114   |
| Sim 10     | 31       | 38       | 69    | 32       | 38       | 70    | 65       | 42       | 107   |
| Sim 11     | 53       | 53       | 106   | 54       | 55       | 109   | 51       | 37       | 88    |
| Sim 12     | 43       | 38       | 81    | 43       | 39       | 82    | 64       | 38       | 102   |
| Sim 13     | 61       | 49       | 110   | 63       | 53       | 116   | 31       | 38       | 69    |
| Sim 14     | 43       | 38       | 81    | 43       | 38       | 81    | 46       | 43       | 89    |
| Sim 15     | 49       | 37       | 86    | 50       | 39       | 89    | 43       | 43       | 86    |
| Sim 16     | 56       | 41       | 97    | 57       | 41       | 98    | 44       | 44       | 88    |
| Sim 17     | 34       | 36       | 70    | 34       | 36       | 70    | 43       | 38       | 81    |
| Sim 18     | 43       | 38       | 81    | 50       | 96       | 146   | 28       | 42       | 70    |
| Sim 19     | 36       | 38       | 74    | 35       | 39       | 74    | 32       | 56       | 88    |
| Sim 20     | 57       | 66       | 123   | 66       | 64       | 130   | 56       | 47       | 103   |
| Sim 21     | 59       | 55       | 114   | 59       | 58       | 117   | 31       | 38       | 69    |
| Sim 22     | 63       | 41       | 104   | 64       | 44       | 108   | 56       | 41       | 97    |
| Sim 23     | 53       | 53       | 106   | 64       | 46       | 110   | 34       | 36       | 70    |
| Sim 24     | 72       | 43       | 115   | 77       | 89       | 166   | 43       | 57       | 100   |
| Sim 25     | 43       | 38       | 81    | 44       | 62       | 106   | 47       | 38       | 85    |

Figure 3. Plots of minimum distance
In Table 1, it can be seen that the comparison of the total minimum distance between the classical method and the clustering method. The percentage of the total minimum distance that is not better than the classical method is 12% which is only 6 out of 50 cases. The remaining 88% of the total minimum distance is better by using the clustering method.

### Table 2. Result of t-test

| t-test                   | Value   |
|-------------------------|---------|
| Difference              | -14.240 |
| t(observd value)        | -4.861  |
| t(critical value)       | -1.677  |
| Degree of freedom       | 49      |
| p-value (one-tailed)    | < 0.0001|
| α                       | 0.05    |

After that, a two-means difference test for the minimum total distance between the clustering method and the classic method was carried out. The results of these tests are presented in Table 2. From Table 2, we know that $t_{\text{observed value}}$ or $t_{stat}$ is smaller than $t_{\text{critical value}}$, so based on the decision-making criteria in the statistics test, $H_0$ is rejected, where $H_0: \mu_1 - \mu_2 = 0$. In addition, the acceptance criterion of $H_0$ can also be found on a $p-value$, i.e. the probability value of obtaining a result is at least as extreme as the observed result of statistical hypothesis testing, assuming that the null hypothesis is true. The $p-value$ is used as an alternative rejection point to provide the smallest significance level for rejecting the null hypothesis. A smaller $p-value$ means that there is more substantial evidence supporting the alternative hypothesis. The result of the $t$-test in Table 2 obtained information that the $p-value < 0.0001$, so $H_0$ is rejected.

Based on the result of the $t$-test in Table 2, at the 95% confidence level ($\alpha = 0.05$), there is sufficient evidence to state that the average total minimum distance by the clustering method is significantly smaller than the average total minimum distance by the classical method. Thus, the results of this simulation provide reasons for the importance of using the clustering method in distribution. It can be concluded that the average total minimum distance obtained using the clustering method is smaller than the classic method.

### 4. Conclusion

Making decisions intuitively or classical methods are no longer efficient in solving the problem of items delivery. Optimization methods and their innovations, one of which involves the clustering method, are very influential on the process of items delivery. The collaboration of them results in the total minimum time.

Based on the simulation results and discussion, it can be concluded that the network clustering method applied to the items delivery process results in the total minimum distance. In addition, this network clustering helps group destination locations so that a courier avoids delivering with very far distances between places. On the other hand, optimizing the total delivery distance is more excellent in the items delivery process that does not use the network clustering method. Then, it is possible to find cases of delivering with very far distances between locations by the same courier.

### Acknowledgment

The authors would like to thank the anonymous reviewers for providing constructive comments to improve the earlier version of the manuscript. The first author would like to thank LPDP, through the scheme of BUDI-DN, for providing the scholarship.
References
[1] Taha H A 2016 *Operations research: an introduction (10th edition)* (New Jersey: Pearson).
[2] Yudhanegara M R, Indratno S W, and Sari R R K N 2021 Dynamic items delivery network: prediction and clustering, *Heliyon*, 7 e06934
[3] Anderberg M R 1973 *Cluster analysis for applications* (New York: Academic Press).
[4] Mooi E and Sarstedt M 2011 *A concise guide to market research* (Berlin: Springer).
[5] Yudhanegara M R and Lestari K E 2019 Clustering for multi-dimensional data set: a case study on educational data, *J. Phys.: Conf. Ser.*, 1280 042025.
[6] Yudhanegara M R, Indratno S W, and Sari R R K N 2020 Clustering for items delivery using rule-k-means, *J. Indones. Math. Soc.*, 26 185-191.
[7] Yudhanegara M R, Indratno S W, and Sari R R K N 2020 Clustering for items distribution network, *J. Phys.: Conf. Ser.*, 1496 012019.
[8] Laporte G and Nobert Y 1983 A branch and bound algorithm for the capacitated vehicle routing problem, *OR Spectr.*, 5 77–85.
[9] Newman M E J 2004 Detecting community structure in networks, *Eur. Phys. J. B* 38 321-30.
[10] Yudhanegara M R and Lestari K E 2020 The graduate views on the contributions of higher educations to future careers, *Int. J. Innov. Creativity Chang.*, 14 1140-53.
[11] Pothen A, Simon H D, and Liou K P 1990 Partitioning sparse matrices with eigenvectors of graphs, *SIAM J. Matrix Anal. Appl.*, 11 430-52.
[12] Elsner U 1997 *Graph partitioning* (Chemnits: Technische Universitat Chemnits).
[13] Trivedi S, Pardos Z A, Sarkozy G N, and Heffernan N T 2011 *Proc. Int. Conf. Ed. Data Min.* vol 4 (Eindhoven: Netherlands Organisation for Scientific Research) p138–147.
[14] Lestari K E, Pasaribu, U S, Indratno, S W, and Garminia H 2020 Generating roots of cubic polynomials by Cardano’s approach on correspondence analysis, *Heliyon*, 6 e03998.
[15] Dror M and Trudeau P 1990 Split delivery routing, *Nav. Res. Logist.*, 37 383-402.
[16] Reinelt G 2012 *TSPLIB: A travelling salesman problem library*, *ORSA j, comput. (Online)*, 3 376-84.
[17] Bernardino R and Paias A 2018 Solving the family travelling salesman problem, *Eur. J. Oper. Res.*, 267 453-66.
[18] Gutin G and Punnen A 2004 *The travelling salesman problem and its variations* (Boston: Kluwer Academic Publishers).
[19] Walpole R E, Myers R H, Myers S L, and Ye K 2012 *Probability and statistics for engineers and Scientists (9th edition)* (Boston: Pearson).