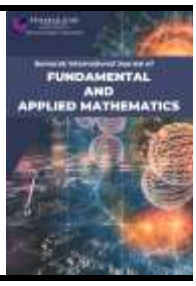
Semarak International Journal of
Fundamental and Applied Mathematics

Journal homepage:
https://semarakilmu.com.my/journals/index.php/sijfam/index
ISSN: 3030-5527



Network Indicators of Public Bus Transportation in Klang Valley and Their Correlations

Nur Umaisara Rashid[1], Chan Kar Tim[2,*], Umair Abdul Halim[3]

[1] Laboratory of Computational Sciences and Mathematical Physics, Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia, 43400, Selangor Malaysia
[2] Department of Physics, Faculty of Science, Universiti Putra Malaysia, 43400 Selangor, Malaysia
[3] Department of Physics, Foundation Studies for Agricultural Sciences, Universiti Putra Malaysia, 43400 Selangor, Malaysia

ARTICLE INFO

Article history:
Received 2 October 2024
Received in revised form 30 October 2024
Accepted 27 November 2024
Available online 15 December 2024

Keywords:
Complex network; Transportation network; Clustering coefficient; Centrality; Robustness

ABSTRACT

Complex network-based analysis of the real-world system has frequently assisted researchers in providing a crucial and extensive understanding of a variety of interconnected useful data especially on public transportation services. In this paper, we construct a public bus transportation network of eight zones in Klang Valley before evaluating their characteristics based on some network measurements such as average degree and clustering coefficient followed by calculation on five network indicators namely accessibility, centralization, robustness, directness, and service connectivity. These network indicators are helpful in providing extensive analysis of the network performance. Other than that, we also analyse the correlation between the network indicators. It is found that the different network indicator is beneficial in describing the network performance based on different network components and characteristics such as how betweenness centrality and closeness centrality are applied as part of the centralization and accessibility measurement respectively.

1. Introduction

The world is abundant with complex systems in various areas, ranging from social and technological systems to information and biological systems [1]. Some examples of complex systems are the Internet [2], the World Wide Web [3], social communities [4], biochemical networks [5], business relationships [6], and public transport [7]. A complex system usually consists of many individual components that interact together collectively. These interactions of the components exhibit certain behaviour which can be detected at the system level. One well-known approach (network theory) is associating the complex system as networks where each element is represented as nodes while the interaction between them is represented as edges. These networks that represent real-world complex systems are called complex networks and they usually display non-trivial topological features [1]. With this simplification, it becomes possible to analyze its properties and

* *Corresponding author.*
E-mail address: chankt@upm.edu.my

https://doi.org/10.37934/sijfam.4.1.1731

gain insights into how the system functions, how it responds to changes and disruptions, and how to design more robust and resilient systems in the future [8].

Public transportation such as bus and light rapid transit (LRT) or mass rapid transit (MRT) is a vital structure of a city. Public transport is often regarded as an indicator of how well urban planning is executed by assessing the residents' commuting experience. Much research work has been conducted under the complex network framework on public transport such as the sub-way [9], bus [10], and metro [11]. The studies are mainly focus on big cities where there is a more suitable land structure [12]. Starting from the work of Tsiotas and Polyzos [13], which talks about the interregional road structure of Greece which highly influenced the placement of public transportation that contributed to the workflow analysis to the work of Wang *et al.,* [14], which highlights the contribution of complex network modeling to identify stations with high capacity at peak time, it shows that complex network approach can be a tremendous factor in increasing the efficiency of public transport service in the near future. In other words, a complex network approach to public transportation networks can provide solutions to the policy maker or urban planner in tackling congestion issues as well as fully utilise the service and improving public transportation [15].

In this paper, we build a network consisting the information on public bus transportation service (PBTS) in eight zones in Klang Valley, Malaysia using a complex network approach based on *l*-space and *p*-space representation. The data for the PBTS can be obtained from their official MyRapid website (myrapid.com.my). Its characteristics are then measured using various network indicators including accessibility, centralization, robustness, service connectivity, and directness. Finally, Pearson's correlation analysis is used to verify the relation between network indicators.

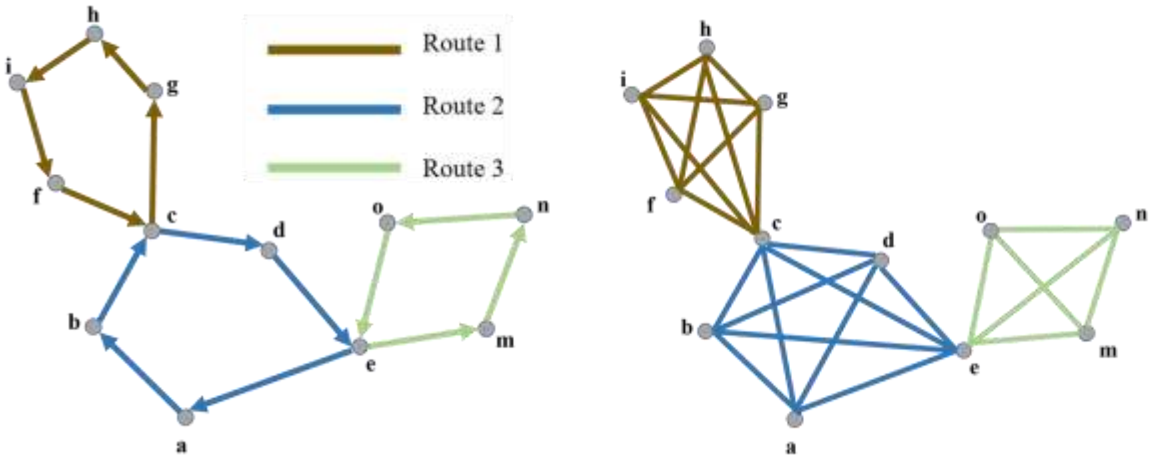
2. Network Representation

A key methodology in the examination of complex networks is to consider the network as a graph ***G (V, E)***, where nodes ***V*** are connected to each other by edges, ***E*** [1]. Nodes that are linked by an edge are referred to as neighbours or adjacent nodes. A graph is said to be complete if all of its nodes are linked to each other. Conversely, a graph is considered disconnected if there is no path that connects all pairs of nodes. When a graph has a path connecting any pair of nodes, it is known as a connected graph.

An undirected network (also called an undirected graph) is a type of network in which the connections (edges) between nodes have no inherent direction. This indicates that if node A is connected to node B, then node B is also connected to node A. In other words, the edges are bidirectional [1]. An example of an undirected network is a social network where the connections between individuals represent friendships or other relationships [16], and the directionality of these connections is not significant.

On the other hand, a directed network (also called a directed graph) is a type of network in which the connections between nodes have a specific direction. This means that if node A is connected to node B, it does not necessarily mean that node B is also connected to node A. In other words, the edges are unidirectional. An example of a directed network is a transportation network where the connections between cities represent one-way roads or air routes, and the directionality of these connections is crucial to determining the fastest or most efficient routes between locations.

In this work, the network is represented in two different network spaces namely *l* -space and *p*-space. In *l* -space, the network represents the structure of the public transportation service where the nodes are the bus stops and the edges connecting the nodes are if there exists at least one route connecting the bus stops [17]. Meanwhile in *p*-space or also known as the transfer network, the nodes are the bus stops and there is an edge between each bus stop or nodes that are servicing the same route [18]. The examples of these two network spaces are shown in Figure 1. By analysing the network in both *l*-space and *p*-space representations, researchers can identify the busiest station while also obtaining information on transfer stations and average number of transfers which is a critical aspect of travel scheduling [19].



(a) (b)

Fig. 1. Examples of a directed PTN in (a) *l*-space, (b) *p* -space

3. Construction of Public Bus Transportation Network (PBTN)

This work focuses on the public bus transportation service (PBTS) in the Klang Valley area of Malaysia, which is the hub of Malaysia's public transportation system. With a population of over 9.0 million [20], this densely populated region is home to numerous offices, including banks, government-linked companies (GLCs), and the headquarters of many international companies. A thorough understanding of the transportation network's mechanism and operation is crucial for improving the efficiency and effectiveness of the system, as well as for preparing transportation providers for future expansions of the service.

The PBTS, also known as RapidKL is provided by the government-funded company, Prasarana Sdn. Bhd. RapidKL bus services started their operation in 2002 and have been expanding since. The data collected for this work includes the bus routes and each stop in the routes. This information is publicly accessible on their official website. The service is available for eight zones in the Klang Valley (see Figure 2) area namely Ampang, Cheras, Damansara, Jalan Pahang, Jalan Ipoh, Jalan Klang Lama, Lebuhraya Persekutuan and Sungai Besi. Figure 3 shows the construction of the network of Klang Valley based on *l*-space and *p*-space network representation while the number of nodes, N, and number of edges, E of the networks are shown in Table 1.

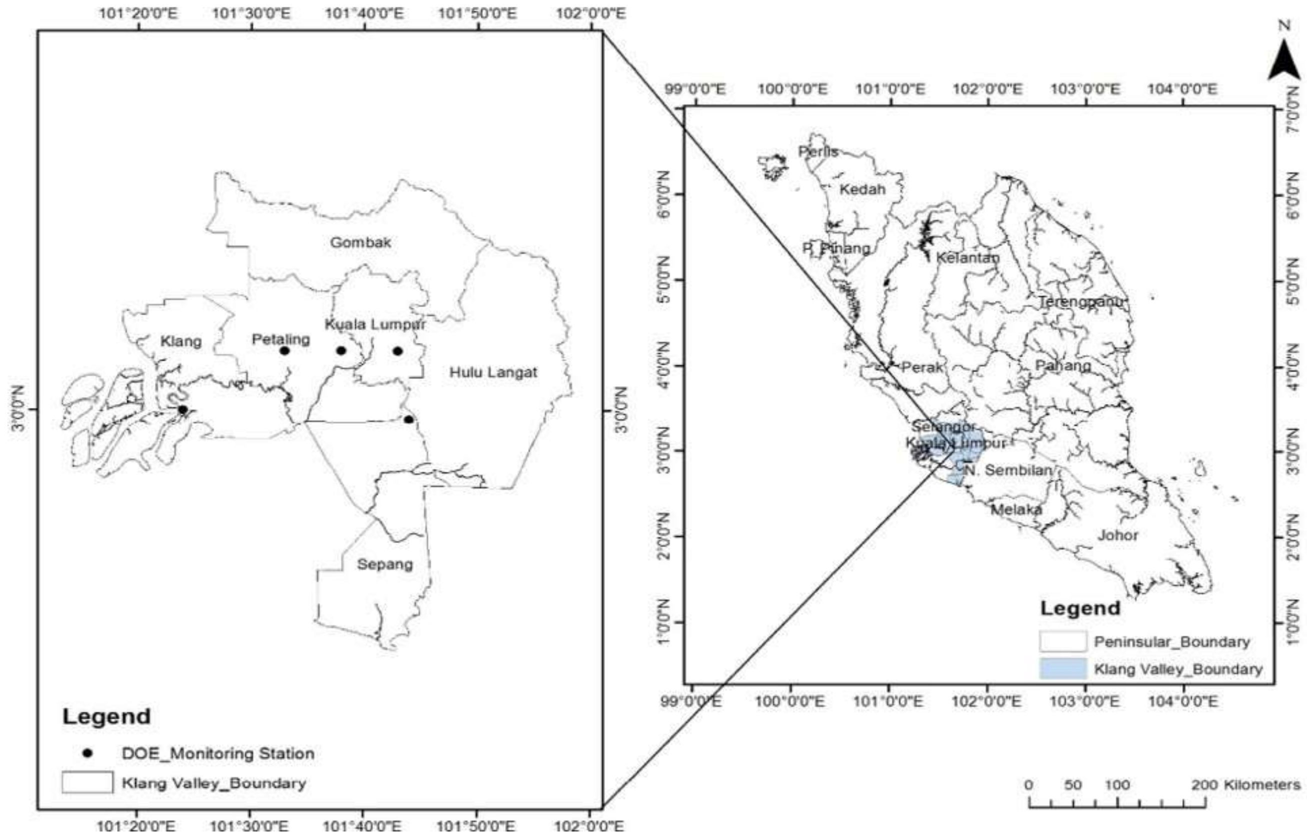


Fig. 2. Map of Peninsular Malaysia and the Klang Valley boundary [21]

(a) (b)

Fig. 3. The combined networks in two different spaces namely (a) *l*-space, (b) *p* -space

Table 1
Measurements of the bus transportation in Klang Valley

Zones	Route	Number of nodes, N	Number of edges in l-space, E_l	Number of edges in p-space, E_p
Klang Valley	150	3565	4423	126047

PBTS has a total of 3565 bus stops and 150 routes. The construction of a public bus transportation network (PBTN) is performed by combining all 150 routes from all zones and as shown in Figure 3. Afterwards, it is observed that the network consists of 10 separate network components with one of them being significantly bigger than the other. It is also observed that 11 routes that are not included in the biggest component. It is analysed that these 11 routes are from the Cheras, Jalan Ipoh, Jalan Klang Lama, Lebuhraya Persekutuan and Sungai Besi. These isolated routes are mainly to cater to the local area. As an example, in the Lebuhraya Persekutuan region, there is one route that is constructed for the area namely route T778 which consists of bus stops such as LRT USJ 21, FAIRVILLE, SMK USJ 23, and ONE CITY. The route is shown on a map in Figure 4.

Fig. 4. Map of the T778 route [22]

The mainly connected network component which consists of 92% of all the routes provided by the PBTS is extracted and labelled as A1 for further analysis on the PBTN. The measurements of the A1 network. Table 2 shows measurements of the bus transportation in Klang Valley.

Table 2
Measurements of the bus transportation in Klang Valley

Zones	Label	Route	Number of nodes, N	Number of edges in l-space, E_l	Number of edges in p-space, E_p
Klang Valley	A1	150	3565	4423	126047

4. Network Analysis of the PBTN

There are a handful of parameters that can be measured from the complex network. some of the common measurements are average path length, vertex degree, clustering coefficient, diameter, and density. These network properties are important as they can give some general observations on the PBTN. To have more details on the PBTN, more specific network indicators such as accessibility, centralization, robustness, directness, and service connectivity are calculated in this work.

PBTN is a directed network because the bus operates on a fixed cycle. For example, one of the routes will start at AJ468 HUB PANDAH INDAH and stop at each 84 bus stops along the route before eventually ending up at the starting bus stop to complete a cycle

4.1 Average Path Length, L and Network Diameter, D

Average path length is the mean of the shortest path between two nodes also known as mean distance. If two nodes are disconnected, meaning there is no path between them, then the path length between them is infinite. If the two disconnected nodes exist in a network, that will also result in the average path length in the network becoming infinite. One way to avoid this problem is to calculate only from nodes in the largest connected component. Mathematically, the average path length, L is shown as:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{1}$$

where where N is the number of nodes and d_{ij} is the distance between two nodes, i and j. From Table 3, it shows that the PBTN has a L_l of 54.09. This value can also be regarded as the number of bus stops one pass by on average when travelling using the PBTS. However, this number is not taking into account the possibility of the need to make a transfer instead only highlighting the shortest path to get from one bus stop to another on average. Meanwhile, for *p*-space network, the path length of an individual node to the other nodes plays a huge role in identifying the number of transfers between nodes. The number of transfers shows the times needed by passengers to change their course of route to reach a certain destination [17]. The formula [10] is as follows:

$$Number\ of\ tranfer = Path\ length - 1 \tag{2}$$

Table 3
Measurements of the parameters for A1

Network	Average degree, $\langle k \rangle_l$	Average path length, L_l	Clustering coefficient, C_l	Density, d_l	Diameter in *l*-space, D_l
A1	1.253	54.09	0.008	0.0003	206

This number of transfers can also be described as the number of times the passengers need to get off at a bus stop to change their route on a different bus. This is crucial in time planning of any trip. Table 4 shows the measurements of average path with the number of transfers.

Table 4
Measurements of average path length in *p*-space, L_p of PBTN and the average number of transfers

Network	L_p	Average number of transfers
A1	4.29	3.29

It is important to note that Lp of PBTN is useful to explain the average number of transfers which for PBTN is 3.29. This signifies that the user needs to transfer to another route to reach their destination on average for 3.29 times.

Beside average path length, network diameter is also relevant in demonstrating the size of the network [1]. Network diameter is the longest distance among all the shortest path lengths calculated. In PTN, network diameter can be regarded as the indicator on the coverage area of the public transportation services where when there are more new bus stops are added especially in a newly developed area.

From Table 3, the network diameter of the PBTN is 206 which shows that there are 206 paths between a pair of bus stops that are the most distant from each other. On the other hand, a large network diameter may be interpreted as a longer travel time, which can contribute to the possibility of delays since there are more bus stops between pair of bus stops which may explain the connectivity of the network.

Apart from that, the connectivity of the entire network can also be evaluated from the number of edges. This usage of the information on the edge count can be utilised in the calculation of network density [1] which reflects the overall tightness of the entire network. In directed PBTN, the network density can be calculated as follows:

$$Network\ density = \frac{Actual\ edges}{Possible\ edges} \tag{3}$$

where the possible edges of a directed network is given as *N(N-1)*. Network density can also be evaluated as having lower direct connections among the bus stops such that users might need to take detours or longer routes to reach their desired destinations. From Table 3, we can conclude that the PBTN has a very low density which signifies the low inter-connectedness. This could be seen as the amount of bus stops are high to cover more area, but there is a minimal bus stop that connects these bus stops from different routes together

4.2 Average Degree, <k> and Clustering Coefficient, C_i^{cyclic}

For each bus stop, there is at least one in-edge and one out-edge. The value stated in Table 3 is the value for both in-degree and out-degree. This small value of the average degree can explain the structure of the network that is sparse. It is also important to pay attention to the individual bus stops that have a high degree. These *l* -space networks consist of directed edges, so we must look at the in-degree, out-degree, and the combination of in-degree and out degree. Other than that, a sparse or dense network can also be observed by evaluating the clustering coefficient. The bus stops with the highest degree are listed in Table 5 with their respective in-degree, out-degree, and clustering coefficients.

Table 5
The first six nodes with the highest degree, *k*

Bus stop	*k*	In-degree	*Out-degree*	C_i^{cyclic}
Hentian Bandar Seksyen 14	11	5	6	~0.001
Tmn Greenwood Jln Batu Caves	10	5	5	~0.001
SJ237 Terminal Puchong Utama	10	5	5	~0.001
Hentian Seksyen 13	10	5	5	0.042
Spg Tmn Melati MRR2	9	5	4	0.059
PJ84 PJ Walk	9	6	3	0.059

A throughout check on the network found that the bus stops that have highest in-degree is not necessarily have the highest out-degree as can be seen in Table 5 for bus stop “PJ84 PJ WALK”. This is because some of th bus stops servicing for more than one particular route. During the construction of the network, any redundant or similar edges are deleted. These edges are when two bus stops servicing at least two similar routes which resulting two out-degrees or two in-degrees between the same pair of bus stops. Bus stops with a huge difference in their number of in-degree and out-degree will influence the number of passengers at the bus stop at one time. For example, bus stop “PJ84 PJ WALK” has 6 in-degrees, but 3 out-degrees, there will be many passengers dropped at bus stop “PJ84 PJ WALK” but fewer buses to pick up the passengers hence causing the bus stop to be busy. The local authorities and the service provider can work together to figure out the suitable maintenance needed at the bus stops to ensure a smooth traveling experience for the users by providing a larger waiting area.

Meanwhile, another measurement that can be crucial in the analysis of the PBTN structure is the clustering coefficient which emphasizes the measure of the degree to which bus stops in a PTN tend to cluster together and is given as:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{4}$$

where the clustering coefficient for node *i*, *Ci* is expressed as:

$$C_i = \frac{number\ of\ triangles\ connected\ to\ node\ i}{number\ of\ triples\ centeres\ around\ node\ i} \tag{5}$$

It is important to point out that the clustering coefficient for the directed network is different from the undirected network. In a directed network, there is an in-degree clustering coefficient and an out-degree clustering coefficient. Fagiolo [23] did some classification on the triangles such as the node is involved in a cyclic, acted as a middleman, or there are also cases where there are two in-edges or two out-edges toward the bus stop in the triangle among the other two bus stop and is expressed as:

$$C_i^{cyclic} = \frac{number\ directed\ cyclic\ triangle\ with\ i}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}} \tag{6}$$

where d_i^{in} is the number of in-degree of node *i*, d_i^{out} is the number of out-degree of node *i*, and $d_i^{\leftrightarrow}$ is the number of a pair of edges a same pair that includes node *i*. In the PBTN of Klang Valley, the clustering coefficient is very low at 0.008. This value shows that the bus stops in the network are sparsely clustered to each other. We identify a few bus stops with the highest clustering coefficient which are shown in Table 6. From Table 6, the bus stops, with the highest clustering coefficient do not necessarily have the highest degree, and in Table 5, the bus stops with the highest degree do not necessarily have the highest clustering coefficient. This happens when certain major bus stops act as hubs with high degrees but lower clustering coefficients in which long-distance connections are highlighted as the main target for the users and efficient travel between major hubs, resulting in lower local clustering [24].

Table 6
Measurements of C_i^{cyclic} clustering coefficient and degree, *k* of the PBTN

Bus stop	C_i^{cyclic}	k
Z&R Rest Jln Wangsa Melawati 3	1	2
Spg Sunway Kayangan	1	2
SL3111 LRT Gombak	1	2
Sblum Spg Kg Subang	1	2
SA247 Pusat Komersial Seksyen 13	1	2
Rumah 148 Jln Ayer Panas	1	2
Rest Jamrut	1	2
Pusat Methodist MRR2	1	2
Petronas Jln Makmur	1	2
Pasngsapuri Sri Puteri Blk A	1	2
KL2080 PPR Sungai Bonos	1	2
Kedai Bunga Jln Kg Bdr Dalam	1	2
Flat Ukay Perdana	1	2
Flat Sg Bonos	1	2

4.3 Specific Network Indicators for PBTN

Relevant characteristic for PBTN is measured using five specific network indicators namely accessibility, centralization, robustness, directness, and service connectivity. The first three indicators are measured in *l*-space while the last two indicators and calculated using *p*-space. The results of all network indicators for the networks are shown in Table 7.

Table 7
Measurement of network indicators in all networks

Bus stop	Network	Accessibility, A	Centralization, *C*	Robustness, R	Directness, D	Service Connectivity, SC
Ampang	R1	0.037	0.709	0.131	0.509	0.883
Cheras	R2	0.022	0.907	0.056	0.538	0.935
Damansara	R3	0.050	0.689	0.065	0.600	0.927
Jalan Pahang	R4	0.025	0.942	0.117	0.501	0.868
Jalan Ipoh	R5	0.028	0.808	0.118	0.622	0.902
Jalan Klang Lama	R6	0.031	0.896	0.103	0.499	0.905
Lebuhraya Persekutuan	R7	0.029	0.567	0.128	0.410	0.865
Sungai Besi	R8	0.032	0.877	0.049	0.568	0.954
Klang Valley	A1	0.019	0.643	0.127	0.282	0.872

4.3.1 Accessibility

The quality of mobility and connectivity in Klang Valley and its respective zone can be measured using accessibility which can be referred to as the ease of travelling from one station to another station in PBTN. This accessibility value ranges from 0 to 1 with 1 being the network with the highest accessibility. Accessibility measures consider the closeness centrality of the network and the number of nodes/stations and the formula adapted from [25] is as follows:

$$Accessibility = \frac{Closeness\ Centrality}{N} \tag{7}$$

where closeness centrality considers the distance between a pair of nodes [26] expressed as:

$$Closeness\ centrality = \frac{N-1}{\sum_j d_{ij}} \tag{8}$$

This centrality value considers the distance between each pair of bus stops, with pair of bus stops with a smaller distance having higher closeness centrality which later indicates higher accessibility. In a way, we evaluate accessibility by looking at the mean closeness centrality overall bus stops. From Table 7, R3 has the highest accessibility with the value a of 0.05 compared to other networks. This value can somewhat be related to the low average shortest path length. Meanwhile, the accessibility value for the A1 network is much lower which suggests that the bus stops in the network are not fully connected. This indicated that more time or path is needed to reach any node. In the transportation system. Accessibility is often measured using metrics such as travel time, distance, or cost which reflects the time and effort required to travel between locations.

4.3.2 Centralization

Centralization mainly covers the centrality of the nodes emphasizing the involvement of each node in the path between two other nodes. In this con- text, we utilize the betweenness centrality. Betweenness centrality is one of the centrality measures that focuses on the node components of the network especially those nodes that are included between the pair of any pair of nodes. Betweenness centrality is particularly used for centralization measures because it captures the node's role as a mediator or a connector between a pair of nodes alongside their path. Nodes that have high betweenness centrality are the nodes that are crucial to make sure the shortest path of any two nodes is at the optimal level. Betweenness centrality is calculated as follows:

$$Betweeness\ centrality = \frac{\sum n_{ij}^{v}}{n_{ij}} \tag{9}$$

where n_{ij} is the number of shortest paths from *i* to j passing through *v* and n_{ij} is the number of paths from *i* to j. Meanwhile, the equation on centralization which utilizes the idea of the betweenness centrality [27] is expressed as:

$$Centralization = \frac{\sum(maxC_B(v) - C_B(v))}{N^3 - 4N^2 + 5N - 2} \tag{10}$$

Centralization in PTN can help describe the node components that are busier than others. For this work, the centralization of the whole network is considered. The range of this centralization measurement is between 0 to 1 with 1 being the network with very strong node centralization.

All of the networks have more than 0.5 value for centralization. This suggests the existence of a hub or center area in the network. In PTN, having high centralization is key to recognizing and later improving any particular stop area such as for facility maintenance and security surveillance. High centralization is also beneficial for the users of public transport as it suggests that some stations or stops are connected to many other stations. Among all eight networks, R4 has the highest value of centralization indicator. Since centralization is highly related to the betweenness centrality, we took a quick look into the individual betweenness centrality of each node in the R4 network. The highest betweenness centrality is the node "KL68 MONORAIL CHOWKIT". This node is located in between Kampung Baru and Chow Kit which is a prominent area of the Kuala Lumpur city among the locals.

Meanwhile, from the A1 network, the node with the highest betweenness centrality value is located in the R3 and R5 namely node "KL108 KOTA RAYA". This bus stop is located in the heart of Kuala Lumpur city with many popular spots in the area such as Petaling Street Market and Central Market.

4.3.3 Robustness

PTN's robustness depends on its structure and preparedness for bus stops or route failures, which can be managed through redundant routes. The robustness is defined as follows:

$$Robustness = \frac{E - N + 1}{(2N) - 5} \tag{11}$$

The robustness measurements are based on the redundant path that exists between two pairs of bus stops. The A1 network also has a small robustness value of 0.127. This emphasizes that if there are failures in the network or road disturbances, it will be difficult for the bus service to provide better recommendations on alternative routes that might have been available to the users. This will eventually cause inconveniences and a bad user experience such as a prolonged delay time.

Overall, the robustness of all the networks in Klang Valley is relatively low. From Table 7, network R1 recorded 0.131 which is the largest robustness among all the other networks

4.3.4 Directness

Public transport services thrive to become an efficient service by minimizing the cost and time taken to reach a destination from any starting point. From the network viewpoint, directness is related to the distance between a pair of bus stops and can be expressed as:

$$Directness = \frac{1}{N(N-1)} d_{ij}^{p} \quad (12)$$

where d_{ij}^{p} is the distance in *p*-space between node *i* and *j*. In p-space, there is an edge between each of the nodes in the same route. From the network viewpoint, efficiency is related to the distance between a pair of nodes. Public transport services thrive to become an efficient service by minimizing the cost and time taken to reach a destination from any starting point. In this work, the network with the lowest directness is R7 with a 0.41 mean while other networks have more than 0.5 directness value.

The directness of a PBTN demonstrates how direct is the stops with each other. In public transport service, it is preferable if the network can be reached easily with less path, or less transfer needed. One might speculate that the directness is contradictory to the concept of robustness that utilises the number of redundant routes as a benchmark of better PBTN performance but the directness in this context is applied to the *p*-space of the network. This is to observe the transfer needed between the nodes. For nodes that are servicing the same route, the node is direct as there is an edge between them in *p*-space. In the case of a pair of nodes that are not in the same route, the distance between them may vary according to the possibility of a transfer node in between the pair of nodes. The number of nodes, edges, and average shortest path length in *p*-space for all networks are shown in Table 8.

Table 8
Measurement of N, E_p and L_p in all networks

Bus stop	Network	Number of nodes, N	Number of edges on p-space E_p	Average path length in *p*-space, L_p
Ampang	R1	339	9331	2.32
Cheras	R2	344	23871	2.28
Damansara	R3	102	1449	1.96
Jalan Pahang	R4	604	26897	2.32
Jalan Ipoh	R5	349	18406	1.88
Jalan Klang Lama	R6	382	10742	2.33
Lebuhraya Persekutuan	R7	1049	37962	2.83
Sungai Besi	R8	250	8174	2.16
Klang Valley	A1	3262	126074	4.29

4.3.5 Service Connectivity

Service connectivity is one of the network indicators that are useful in interpreting the public transportation service performance highlighting the linkage between any two routes. The service connectivity is expressed as:

$$Service\ Connectivity = \frac{1}{N}\sum_{i=1}^{N} C_i^p \tag{13}$$

where C_i^p is the clustering coefficient in the *p*-space network for node *i*.

Observing the service connectivity value of the A1 network, it is clear that the routes in this PBTN are well connected and the service connectivity of all the public bus transportation in eight networks in Klang Valley is very high as stated in Table 7. This suggests that the bus routes in these services are well connected and as long as the transfer points are identified, users can travel across the network with little to no issue. Having high service connectivity is very important because the connections of the service lines across several areas are key to a well-constructed public bus transportation service.

4.4 Trends of Network Indicator

Descriptive statistics of the network have been calculated in order to study the trend of the network indicators for PBTN. From Table 9 it is observed that centralization and service connectivity have the highest mean value. This shows that the PBTN in Klang Valley has a recognizable hub location and the routes are greatly connected. It is also clear that centralization and robustness have a range of more than 0.5. This could suggest that there is high variability of centralization and robustness indicators among the PBTN in Klang Valley. Meanwhile, a low standard deviation value in PBTN in Klang Valley indicates more consistent and stable performance, with less variability among the data points. It implies a reliable and standardized transportation service across the entire network.

Table 9
Basic statistics of the five indicators; A, C, R, D and SC of all eight regions

Measurements	*A*	*C*	*R*	*D*	*SC*
Mean	0.032	0.724	0.169	0.531	0.905
Median	0.030	0.759	0.118	0.524	0.904
Standard Deviation	0.009	0.214	0.197	0.067	0.032
Min	0.022	0.296	0.049	0.410	0.865
Max	0.050	0.942	0.650	0.622	0.954
Range	0.028	0.646	0.602	0.212	0.089

4.4.1 Pearson's Correlation

Correlation analysis is one of the analyses for studying the relation between two measurements or two sets of data [28]. The analysis is helpful to provide an extensive observation on the characteristics of the network. Pearson's correlation is one of the correlation analyses that is quantified by introducing the correlation coefficient, r_{ab} and the calculation is as follows:

$$r_{ab} = \frac{\sum_{m=1}^{M}(a_m - a)(b_m - b)}{\sqrt{\sum_{m=1}^{M}(a_m - \bar{a})^2 \sum_{m=1}^{N}(b_m - \bar{b})^2}} \tag{14}$$

where a_m and b_m is the two sets of data for each element m with M being the total number of m. This coefficient will have values between -1 to 1. The correlation is considered to be very strong if the Pearson's correlation is more than 0.81 while not correlate if the value is less than 0.20.

The correlation between the network indicators is essential in this work to provide a substantial observation because some of the indicators are influenced by the network components which are nodes and edges. Pearson's correlation values for the network indicators and the network size are shown in Table 10.

Table 10
Basic statistics of the five indicators; A, C, R, D, and SC of all eight regions

Indicators	*A*	*C*	*R*	*D*	*SC*	*N*	*E*
A							
C	-0.27						
R	0.87	-0.12					
D	0.32	0.38	0.35				
SC	0.19	0.22	0.13	0.64			
N	-0.18	-0.47	-0.38	-0.74	-0.84		
E	-0.19	-0.44	-0.36	-0.76	-0.84	0.99	

From Table 10, it is apparent that the number of bus stops and connecting routes negatively impact the service connectivity of the PBTN as the network size increases. Accessibility and robustness, which represent backup route plans and redundant routes, show a strong positive correlation. However, the correlation does not imply causation between the parameters, only their relationship.

5. Conclusions

In this work, the basic characteristics of the public bus transportation in Klang Valley in eight zones are calculated alongside five network indicators namely accessibility, centralization, robustness, directness, and service connectivity. The network indicators are evaluated mainly to emphasize the performance of the network. Some of the network indicators are utilised on *l*-space which are accessibility, centralization, and robustness. Meanwhile, directness and service connectivity are calculated on *p*-space. This is because different network indicators are useful in elaborating different aspects of network characteristics and performance. The evaluation of the network accessibility of all eight zones shows that the PBTNs are poorly accessible from any node in the network. However, the centralization of these networks shows an impressive value where all eight networks have a high centralization with a mean of 0.724. The robustness of the networks recorded an average value of 0.169 169 indicating the difficulty of users to navigate around Klang Valley in the case of service disruptions. Apart from that, the directness shows a moderate mean value of 0.531 meanwhile the service connectivity has the highest mean value of 0.905. These two indicators are useful in emphasizing the transfer property and the route connections in the PBTN. Meanwhile, accessibility and robustness showed the strongest correlation which is highlighted by the interconnection of the public bus transportation service alongside the redundancy and alternative paths provided by the service.

Acknowledgement
The authors wish to acknowledge the financial support provided by Universiti Putra Malaysia under UPM Putra Grant with project number GP-IPS/2022/9726900.

References

[1] Newman, MARK EJ. "Networks: an introduction." (2010). https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

[2] Ravasz, Erzsébet, and Albert-László Barabási. "Hierarchical organization in complex networks." *Physical review E* 67, no. 2 (2003): 026112. https://doi.org/10.1103/PhysRevE.67.026112

[3] Barabási, Albert-László, Réka Albert, and Hawoong Jeong. "Scale-free characteristics of random networks: the topology of the world-wide web." *Physica A: statistical mechanics and its applications* 281, no. 1-4 (2000): 69-77. https://doi.org/10.1016/S0378-4371(00)00018-2

[4] González, Marta C., Hans J. Herrmann, J. Kertész, and Tamás Vicsek. "Community structure and ethnic preferences in school friendship networks." *Physica A: Statistical mechanics and its applications* 379, no. 1 (2007): 307-316. https://doi.org/10.1016/j.physa.2007.01.002

[5] Barkai, Naama, and Stan Leibler. "Robustness in simple biochemical networks." *Nature* 387, no. 6636 (1997): 913-917. https://doi.org/10.1038/43199

[6] Kawamoto, Hirokazu, Hideki Takayasu, and Misako Takayasu. "Analysis of network robustness for a japanese business relation network by percolation simulation." In *Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014*, pp. 119-127. Springer International Publishing, 2015. https://doi.org/10.1007/978-3-319-20591-5_11

[7] Angeloudis, Panagiotis, and David Fisk. "Large subway systems as complex networks." *Physica A: Statistical mechanics and its applications* 367 (2006): 553-558. https://doi.org/10.1016/j.physa.2005.11.007

[8] Derrible, Sybil, and Christopher Kennedy. "The complexity and robustness of metro networks." *Physica A: Statistical Mechanics and its Applications* 389, no. 17 (2010): 3678-3691. https://doi.org/10.1016/j.physa.2010.04.008

[9] Zhang, Jianhua, Xiaoming Xu, Liu Hong, Shuliang Wang, and Qi Fei. "Networked analysis of the Shanghai subway network, in China." *Physica A: Statistical Mechanics and its Applications* 390, no. 23-24 (2011): 4562-4570. https://doi.org/10.1016/j.physa.2011.06.022

[10] De Bona, A. A., K. V. O. Fonseca, M. O. Rosa, R. Lüders, and M. R. B. S. Delgado. "Analysis of Public Bus Transportation of a Brazilian City Based on the Theory of Complex Networks Using the P-Space." *Mathematical Problems in Engineering* 2016, no. 1 (2016): 3898762. https://doi.org/10.1155/2016/3898762

[11] Kanwar, Kushal, Harish Kumar, and Sakshi Kaushal. "Complex network based comparative analysis of Delhi Metro network and its extension." *Physica A: Statistical Mechanics and its Applications* 526 (2019): 120991. https://doi.org/10.1016/j.physa.2019.04.227

[12] Ding, Rui, Norsidah Ujang, Hussain bin Hamid, and Jianjun Wu. "Complex network theory applied to the growth of Kuala Lumpur's public urban rail transit network." *PloS one* 10, no. 10 (2015): e0139961. https://doi.org/10.1371/journal.pone.0139961

[13] Tsiotas, Dimitrios, and Serafeim Polyzos. "Introducing a new centrality measure from the transportation network analysis in Greece." *Annals of Operations Research* 227 (2015): 93-117. https://doi.org/10.1007/s10479-013-1434-0

[14] Wang, Junjie, Yishuai Li, Jingyu Liu, Kun He, and Pu Wang. "Vulnerability analysis and passenger source prediction in urban rail transit networks." *PloS one* 8, no. 11 (2013): e80178. https://doi.org/10.1371/journal.pone.0080178

[15] Derrible, Sybil, and Christopher Kennedy. "Characterizing metro networks: state, form, and structure." *Transportation* 37 (2010): 275-297. https://doi.org/10.1007/s11116-009-9227-7

[16] Hollstein, Betina, Tom Töpfer, and Jürgen Pfeffer. "Collecting egocentric network data with visual tools: A comparative study." *Network Science* 8, no. 2 (2020): 223-250. https://doi.org/10.1017/nws.2020.4

[17] Von Ferber, Christian, Taras Holovatch, Yu Holovatch, and Vasyl Palchykov. "Public transport networks: empirical analysis and modeling." *The European Physical Journal B* 68 (2009): 261-275. https://doi.org/10.1140/epjb/e2009-00090-x

[18] Sen, Parongama, Subinay Dasgupta, Arnab Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. "Small-world properties of the Indian railway network." *Physical Review E* 67, no. 3 (2003): 036106. https://doi.org/10.1103/PhysRevE.67.036106

[19] Zhang, Yihan, Qingnian Zhang, and Jigang Qiao. "Analysis of Guangzhou metro network based on L-space and P-space using complex network." In *2014 22nd international conference on geoinformatics*, pp. 1-6. IEEE, 2014. https://doi.org/10.1109/GEOINFORMATICS.2014.6950806

[20] Current Population Estimates, Department of Statistics, Malaysia, 2022.

[21] Mohd Shafie, Siti Haslina, Mastura Mahmud, Suzani Mohamad, Nor Lita Fadilah Rameli, Ramdzani Abdullah, and Ahmad Fariz Mohamed. "Influence of urban air pollution on the population in the Klang Valley, Malaysia: a spatial approach." *Ecological Processes* 11, no. 1 (2022): 3. https://doi.org/10.1186/s13717-021-00342-0

[22] Station, RailTravel. RapidKL Service T7778 from USJ 21 LRT to ONE CITY USJ 23 by bus, 2016.

[23] Fagiolo, Giorgio. "Clustering in complex directed networks." *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 76, no. 2 (2007): 026107. https://doi.org/10.1103/PhysRevE.76.026107

[24] Hosapujari, Amalingayya B., and Ashish Verma. "Development of a hub and spoke model for bus transit route network design." *Procedia-Social and Behavioral Sciences* 104 (2013): 835-844. https://doi.org/10.1016/j.sbspro.2013.11.178

[25] Tundulyasaree, Krissada. "Topological characterizing and clustering of public transport networks." (2019).

[26] Cohen, Edith, Daniel Delling, Thomas Pajor, and Renato F. Werneck. "Computing classic closeness centrality, at scale." In *Proceedings of the second ACM conference on Online social networks*, pp. 37-50. 2014. https://doi.org/10.1145/2660460.2660465

[27] Freeman, Linton C. "Centrality in social networks: Conceptual clarification." *Social network: critical concepts in sociology. Londres: Routledge* 1 (2002): 238-263.

[28] Jawlik, Andrew A. *Statistics from a to z: Confusing concepts clarified*. John Wiley & Sons, 2016. https://doi.org/10.1002/9781119272021