https://doi.org/10.1016/j.physa.2018.02.111

This is the Pre-Published Version.

Spatial Analysis of Bus Transport Networks using Network Theory

Shanmukhappa Tanuja[a], Ivan W. H. Ho[a,b], and Chi Kong Tse[a]

[a] *The Hong Kong Polytechnic University, Hong Kong, China*
[b] *The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen,China*

Abstract

In this paper, we analyze the bus transport network (BTN) structure considering the spatial embedding of the network for three cities, namely, Hong Kong (HK), London (LD), and Bengaluru (BL). We propose a novel approach called supernode graph structuring for modeling the bus transport network.A static demand estimation procedure is proposed to assign the node weights by considering the points of interests (POIs) and the population distribution in the city over various localized zones. In addition, the end-to-end delay is proposed as a parameter to measure the topological efficiency of the bus networks instead of the shortest distance measure used in previous works. With the aid of supernode graph representation, important network parameters are analyzed for the directed, weighted and geo-referenced bus transport networks. It is observed that the supernode concept has significant advantage in analyzing the inherent topological behavior. For instance, the scale-free and small-world behavior becomes evident with supernode representation as compared to conventional or regular graph representation for the Hong Kong network. Significant improvement in clustering, reduction in path length, and increase in centrality values are observed in all the three networks with supernode representation. The correlation between topologically central nodes and the geographically central nodes reveals the interesting fact that the proposed static demand estimation method for assigning node weights aids in better identifying the geographically significant nodes in the network. The impact of these geographically significant nodes on the local traffic behavior is demonstrated by a simple simulation using the SUMO (Simulation of Urban Mobility) tool which is also supported by real-world empirical data, and our results indicate that the traffic speed around a particular bus stop can reach a jammed state from a free flow state due to the presence of these geographically important nodes. A comparison of the simulation and the empirical data provides useful information on how the bus operators can better plan their routes and deploy stops considering the geographically significant nodes.

Keywords: Bus tranport network, spatial analysis, supernode structure, node weight, complex networks, static demand estimation

1. Introduction

To decipher the dynamics of real-world complex systems, understanding the underlying connectivity and interaction of the network components is a prerequisite. The inherent behaviors of complex networks are partly encoded in their topological properties which can be analyzed using the concepts recently established in network science, which is a combination of graph theory, statistical physics, control theory and information theory [1]. This research work focuses on the graph theory and statistical physics for analyzing the bus transport network of the three cities namely, Hong Kong, London and Bengaluru.

In a bus transport network, a node represents a bus stop and an edge represents the route between two stops, commonly known as L-Space modeling in complex networks. The key contributions of this paper are three-fold:

1. Unlike the conventional graph representation, our approach uses a novel supernode graph structuring procedure to combine geographically closely associated nodes based on a specific criterion, resulting in a more compact representation. This re-structuring of the network has significant impact on the analysis of different network parameters and the information that can be obtained about the actual network behavior.
2. In graph theory, it is a common practice to assign edge weights to generate a weighted network. However, not much information has been obtained from accurately weighting the nodes [2, 3]. In assigning node weights, a static demand estimation method is proposed by considering the points of interests (POIs) and the population distribution over localized zones in the city. Depending on the POI density and the node occupying index

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

(number of people accessing a particular bus stop) over localized zones, a node weight is assigned to every node in the network.

3. The statistical analysis of various network parameters for the three cities is accomplished to verify the topological behavior. Lastly, the end-to-end travel delay is proposed as a measure to evaluate topological efficiency instead of the conventionally used measure of geodesic distance between nodes. The end-to-end delay method is validated by simulations using the SUMO tool, which also demonstrates the effect of geographically significant nodes over the local traffic behavior. The simulation results are compared with the real-world data to verify the effect of geographically significant nodes and their distance of separation on the local traffic movement.

All the BTN structures analyzed in our study are represented as a single network assuming that a single operator is operating the routes in the network, rather than considering individual operators and dividing the network into subnetworks. Every station is identified by a unique node ID, since some stops have the same name in the network. The graph representation includes both inbound and outbound routes in all the three cities, which makes the graph directed. Our choice of the three cities is based on the geographical landscape, the servicing area, passenger carrying capacity per day and the contribution of the bus transport to the overall public transport. Table 1 shows the statistical details of the bus transport networks for the three cities [4–7].

Modelling large real-world networks (e.g., WWW, protein network, transportation network) as graphs and analyzing their behavior from a network perspective facilitates better understanding of the global and local properties of the network. Specifically, the study of the internet or the World Wide Web played a major role in the development of numerous fundamental properties in network theory. In 1958, Erdos and Renyi [8] proposed the concept of random graphs where the degree distribution follows a Poisson distribution. But the resonating work done by Barabasi [9] posed a challenging question about the actual connectivity of real-world networks, i.e., do many of our real-world networks function seamlessly if they are wired randomly together? This led to the development of scale-free network property which shows a significant deviation in the degree distribution from Poisson to power law. Other significant contributions in the field are credited to Watts and Strogatz [10] for their demonstration of the small world property, Burt [11] for working on missing links in the networks called structural holes, Katz [12], Bonacich [13], Page [14], Kleinberg [15] for their contributions to different centrality measures, Latora and Marchiori [16] for proposing the topological efficiency measure.

Sienkiewicz and Holyst [17] discussed about the statistical analysis of 22 public transport networks in Poland with network size varying from 152 to 2881, which is considered one of the earliest, yet exhaustive studies on application of network theory concepts to public transport networks (PTN). The different cities considered are modeled in L-Space and P-Space. The study reveals that L-Space degree distribution follows a power law, whereas the P-Space structure exhibits exponential behavior. The path length distributions follow an asymmetric unimodal function and all the networks considered show small world property. Chen *et al.* [18] investigated urban bus transport network of four major cities in China: Hangzhou, Nanjing, Beijing and Shanghai. Their empirical results show that the degree distribution in these four cities has an exponential form. They further investigated the number of bus routes that pass a stop and the number of stops per route, denoted by R and S respectively. The values of R and S are used to determine the evolution of the bus transport network. Specifically, a generic model is proposed for the evolution of BTN using the values of R and S which fit their empirical results. Xu *et al.* [19] considered three cities in China and evaluated the scaling laws and the correlation properties using P-Space modeling, where every route is represented as a node, and if a few routes share a common set of nodes between them, they are connected by an edge, and the number of routes sharing these nodes is assigned as an edge weight. The degree distribution of such a weighted network of shared node representation is evaluated and found to follow a power law distribution.

Ferber *et al.* [20] accomplished an exhaustive study of bus transport network by analyzing fourteen major cities across the world with varied network sizes. They generated simple graphs for the network maps, bipartite graphs for routes and stations, and one mode projection of the latter to investigate the inter-relations and spatial embedding properties of routes. Based on empirical observations, they formulated a model with simple growth rules to verify the scale-free behavior of the network. Chatterjee *et al.* [21] performed a detailed statistical analysis of bus network for six major cities in India. They evaluated various network parameters from degree to network robustness by modeling the network in L-Space. The weighted degree distribution of all the networks followed a heavy tailed distribution, the centralities followed a random attachment showing that the network may not be fully scale-free. The networks showed the small world property and though are robust and resilient to random attacks, they are degree sensitive. This work also discussed a concept similar to our proposed supernode model, called the short-distance station pairs, which combines stops that are geographically close to each other with no direct bus

connectivity. Hui *et al.* [22] analyzed the BTN in Beijing where the basic nodal parameter analysis was carried out in L-Space and all the transfer properties were analyzed in P-Space. The degree distribution of Beijing follows a shifted power law for degree > 2, indicating that the network is scale free. The network also behaves as a sparse network because of its very low clustering coefficient value, and shows strong assortative correlation, indicating that higher degree nodes are well connected with other higher degree nodes. A weighted complex network analysis of the travel routes in Singapore was done by Soh *et al.* [23], in which the topological and dynamical properties of the graph structure were considered for analyzing the transport network. This was the first paper that discussed the study of PTN based on geographical properties

The rest of the paper is organized into four sections, Section 2 deals with the restructuring of a conventional graph structure into a supernode structure and adding weights to the network. Section 3 describes a detailed statistical analysis of three bus transport structures. Section 4 evaluates the topological efficiency with the aid of SUMO simulation. The last section provides a conclusion and a brief plan of future work.

Table 1: Statistical details of bus transport network for the three cities.

Parameter	**Hong Kong**	**London**	**Bengaluru**
Population (millions)	7.3	8.6	8.4
Area (km^2)	1104	1579	709
Bus stops	4065	20192	5662
Routes	916	685	2040
Fleet size	5900	9200	6100
Daily passengers (millions)	3.8	5.9	4.9
Overall contribution to PTN (%)	33	21	42

2. Graph representations of bus networks

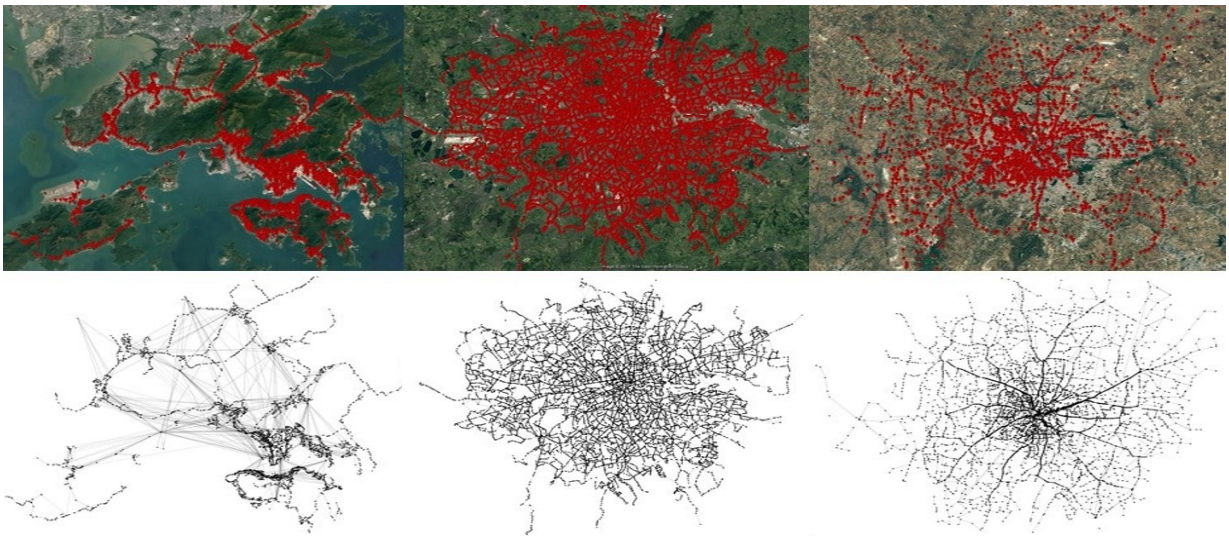
In this section we first discuss the representation of a bus transport network as a standard graph which is later modified with the proposed supernode graph representation. By considering the supernode representation, node and edge weights are then added to the network. We propose a static demand estimation method to assign node weights and the the edge weights are assigned by considering the number of routes operating between the two chosen nodes.

As per conventional graph theory approach, a graph G is a set of V nodes and E edges, i.e., $G = (V, E)$. Considering the spatial analysis of the network, in the current work, a graph G is represented by $G = (V(X, Y), E)$ where V and E are described as:

$$V = \{n_i\,(x_i, y_i) : i = 1, 2, ..., N, x_i = \text{latitude}, y_i = \text{longitude}\} \tag{1}$$

$$E = \{e_{ij} \rightarrow (n_i\,(x_i, y_i)\,, n_j\,(x_j, y_j))\,\forall\,(n_i\,(x_i, y_i)\,, n_j\,(x_j, y_j)) \in V : i = j = 1, 2, ..., N\} \tag{2}$$

where $N = |V|$ indicates the network size. In the subsequent sections, $n_i(x_i,y_i)$ is represented as n_i assuming that a particular node is always identified with its latitude and longitude. The graph is generated in L-Space, i.e., a bus stop represents a node, and there exists an edge between any two nodes if they are consecutive stops on a particular bus route [20]. Multiple edges between the nodes are not considered in a L-Space representation which makes it the representation of actual physical connectivity of the network, i.e., the L-Space representation consists of the bus stops and the presence or absence of connectivity between the stops irrespective of the number of routes between the stops. An $N \times N$ adjacency matrix A with entries a_{ij} is used to describe the connectivity between nodes, where a_{ij}=1 if there exists a route between nodes n_i and n_j, and 0 otherwise. The adjacency matrix is converted to a directed edge list for generating the graph structure. Fig. 1 shows the spatial locations of bus stops and the final bus transport network structure for the three cities. The geographical locations are represented in WGS84 standard using ArcGIS tool [24]. Google Earth is used to visualize the spatial location of the bus stops [25], and Gephi tool is used for graph structure visualization and analysis [26].



(a) (b) (c)

Figure 1: Spatial location of bus stops for the (a) Hong Kong; (b) London; and (c) Bengaluru networks.

2.1. Supernode graph structure

The inspection of spatial embedding of nodes in transport structure analysis has resulted in a new type of network element called *supernode* as shown in Fig. 2a. A supernode is a set of geographically closely associated nodes which satisfy the following condition:

$$d_{ij} < d_{th} \tag{3}$$

where d_{ij} is the geographic distance between two nodes i and j, and d_{th} is a defined threshold distance. If $d_{ij} < d_{th}$, nodes i and j are said to be geographically close to each other. The value of d_{th} is set to be 100 m in this paper assuming that it is a walkable distance to reach any station. The distance between nodes d_{ij} is evaluated using the Haversine formula [27]).

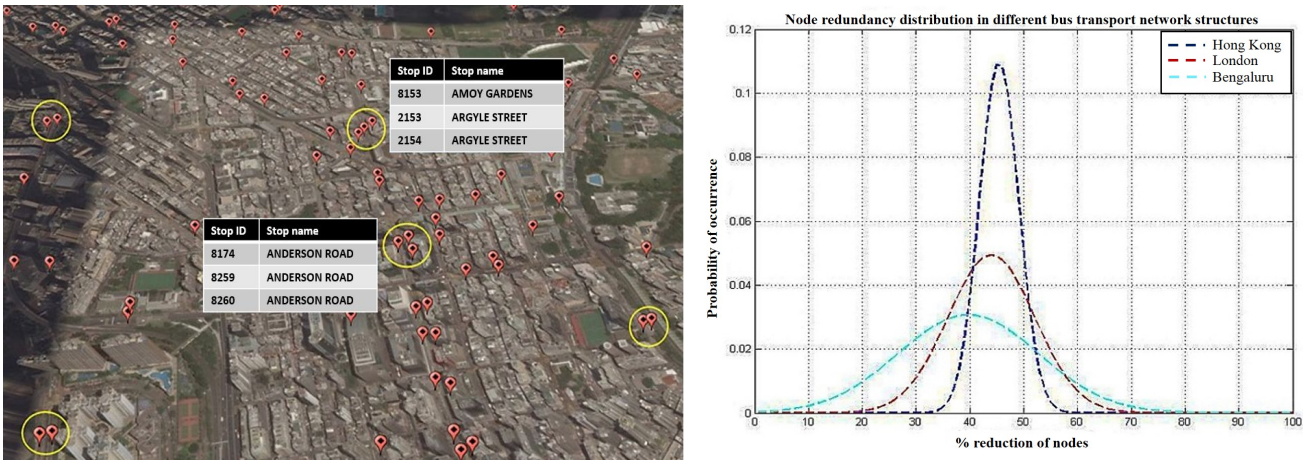
The set of nodes satisfying the above condition are combined to represent a single node called supernode. The combining of nodes is not physical, however, it is a structural re-organization of the network which makes the topological analysis more practical. For example, the bus stops located on either sides of a road segment are treated as a supernode from network analysis perspective. Supernode graph structure can be mathematically defined as

$$\text{SNV} = \left\{ \begin{matrix} sn_j(x_j, y_j), sn_j = \{n_j \forall j = 2, 3, ..., m | (d_{n_i n_j} < d_{th})\}, sn_j(x_j, y_j) = mean(n_j(x_j, y_j)) \\ \exists (sn_p \cap sn_j)_{p \neq j} \neq 0 \end{matrix} \right\} \tag{4}$$

$$\text{SNE} = \{esn_j \rightarrow (sn_i, sn_j) \,|\, (sn_i \ni n_i), (sn_j \ni n_j)\} \tag{5}$$

Thus, a supernode graph structure consists of regular nodes, supernodes, regular edges and superedges. While defining the supernode structure some of the original nodes will be eliminated due to the geographical combining of nodes and there might exist self-loops due to the merging of nearby nodes. Such self-loops are eliminated since they convey little information. The significance of supernode structure can be summarized as follows:

1. Combining the geographically significant nodes in the network improves understanding of the structural behavior of the network. For example, combining nodes with significant degrees in the network helps identify important hubs in the network (discussed in Section 3.6.2)
2. Supernode structure helps to determine the convenient switching points or transfer points which contribute to efficient routing in the network and improve the overall topological efficiency.
3. Supernode structure aids in assigning node weights which reflect the geographical significance of a node (discussed in Section 2.2.1)
4. Supernode structure helps eliminate redundancies in the network for fast computation, i.e., with supernode structure, we can reproduce a network that is close to the original network with a reduced data set.



(a) (b)

Figure 2: (a) An example of geographically closely associated nodes in Hong Kong; (b) redundancy distribution for the three cities.

Table 2: Comparison of network size for the three cities with original and supernode representations.

	Regular structure		**Supernode structure**		**% redundancy**	
	Nodes	**Edges**	**Nodes**	**Edges**	**Nodes**	**Edges**
Hong Kong	4065	11672	2251	8497	45	27
London	20192	24117	11271	21488	44	11
Bengaluru	5662	13266	3724	9832	34	26
Average	-	-	-	-	41	21

Table 2 gives statistical details of the original networks, the supernode networks, and the percentage redundancy for the three networks. The redundancy is calculated over a specific geographical area called zone (ward/district), i.e., redundancy is the total number of nodes present in a particular zone with supernode structure representation as compared to the conventional graph representation. Fig. 2b shows that the percentage redundancy graph follows a normal distribution (with μ equal to 45, 44, 37 and σ equal to 4, 8, 16 for Hong Kong, London and Bengaluru networks, respectively). It is observed that the redundancy in the Hong Kong bus network is maximum as compared to the London and Bengaluru networks. The high density of nodes in certain zones causes the network to shrink more, this is because high density indicates more closely packed nodes, and in the supernode structuring, such close associates are joined as supernodes. Thus, from the percentage redundancy, we can also obtain a better insight into the node distribution pattern of a network. The physical meaning of the redundancy parameter is that with the proposed supernode graph representation method, a network structure close to the original network can be reproduced using the reduced dataset. Thus, although we observe an average of 40% redundancy of nodes in the three cities, only 20% redundancy is observed in the network connectivity (edges), indicating that a major portion of the original network structure is still retained in supernode representation.

2.2. Weighted supernode graph structure

In this section, the supernode structure is considered to model the network as a weighted network by assigning node weight and edge weight. In classical graph theory, assigning weights to edges is a common practice for generating weighted networks, but assigning weights to nodes is less considered [2, 3]. In this work, we propose a static demand estimation method to assign a node weight. The edge weight is represented by considering the number of routes operating between the chosen two nodes.

2.2.1. Node weight

The necessity of a bus stop by the public is greatly influenced by the presence of points of interests (POIs) around the bus stop. The POI can be either a hospital, hotel, office, school, sports arena, cinema hall, shopping

complex or residential apartment. Information about location of POIs around a bus stop provides more practical estimation of the demand serviced by a bus stop. The POIs considered in the current work are broadly divided into four categories, as shown in Table 3. Using these POIs, the node weight can be evaluated by

$$w_i = c_1 \left(\sum_{m=1}^{4} d_m \right)_i + c_2 P_i + c_3 k_i \tag{6}$$

where w_i is the weight of node i, d_m is the number of POIs of category m located around node i within a radius of 100 m, P_i is the total number of passengers accessing node i, k_i is the node degree, c_1, c_2 and c_3 are scaling factors, assumed to be equal to 1 for simplification. In (6), though it is not practical to map certain POIs around a node to specifically one node (since certain POIs can be equidistant to multiple nodes), such shared POIs can be allocated to the nearest node with the least distance. To demonstrate the static demand estimation method for assigning node weight, the Hong Kong bus transport network is taken as an example. However, for the purpose of analysis, we consider POI density and the number of people accessing a bus stop (node occupying index) over localized zones to generate the node weights. Hence, (6) is modified as

$$(w_i)_{\text{zone}} = \left(\frac{c_1 \sum_{m=1}^{4} \rho_m + c_2 \rho_p}{\rho_N} \right)_{\text{zone}} + c_3 k_i \quad ; \text{zone} = 1, 2, ..., n \tag{7}$$

where w_i is the weight of node i in a specific zone. We divide the area into smaller units called zones (districts/wards). Each zone possesses a defined land area, population, POIs and bus stops within it. The term ρ_m denotes the POI density of category m in a zone, ρ_p is the static population density in a zone, ρ_N is the bus stop density in a given zone, k_i is the node degree (discussed in Section 3.1) and the fraction $((\rho_m + \rho_p)/\rho_N)_{\text{zone}}$ denotes the node occupying index. Thus, (7) aids in assigning node weights based on static demand estimated using POIs and population density.The node weight is normalized to ensure the data integrity in all zones, i.e.,

$$(w_{i_\text{norm}})_{\text{zone}} = \left(\frac{w_i - w_{\min}}{w_{\max} - w_{\min}} \right)_{\text{zone}} \quad ; \text{zone} = 1, 2, ..., n \tag{8}$$

Closer the value of w_{i_norm} to one in (8), higher the demand serviced by the node in a zone. Fig. 3a shows a snapshot of the Kowloon zone in Hong Kong with the bus stops and POI locations in the zone. Fig. 3b shows the complete list of bus stops and POI locations in the city considered for validating the static demand method. Fig. 4a shows the nodes (with red color) serving maximum demand in the Hong Kong city based on the normalized node weight calculated using (8) and Fig. 4b shows the heat map indicating the higher demand zones in the city. From (7), we can notice that, node weight is a function of four parameters, i.e., ρ_m, ρ_p, ρ_N and k_i. As ρ_m, ρ_P and k_i increase, w_i also increases. However as discussed in [28], the larger the number of bus stops, the less efficient the network is. Hence, in (7), with w_i being inversely proportional to node density ρ_N, as the number of bus stops in a zone increases, the node weight decreases. As compared to the other three parameters, ρ_p has strong influence on w_i since the growth rate of the population is more drastic as compared to growth rate of the other three parameters. Table 4 demonstrates the method used to assign node weight to a node in a particular zone using the Hong Kong government database. [Note: The demand estimation model is verified only for the Hong Kong network currently, and the POI data considered for demand modeling are obtained from the Hong Kong government database [29], which currently has a few categories not freely available, and hence the data list is not complete].

Table 3: Different POI categories considered for static demand estimation method.

Recreation	**Emergency**	**Education**	**Transportation**
Parks, restaurants, performing venues, libraries and museums, sports grounds, tourist attractions	Hospitals, banks, post-offices, police-stations	Schools and universities, workplaces	Bus stops, taxi-stands ferry service, metro stations and trams

Table 4: Illustration of node weight analysis used to assign a normalized weight (w_{i_norm}) to a node (i) in a particular zone (Central and Western district).

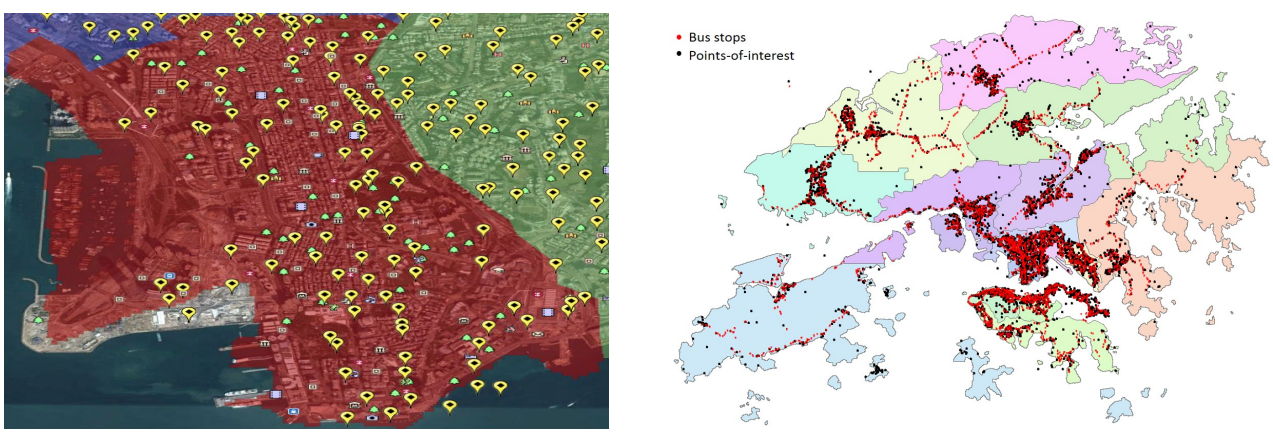
Zone	**Area** (km^2)	ρ_m	ρ_P	ρ_N	$\frac{\rho_m+\rho_P}{\rho_N}$	**Node**(i)	k_i	$\frac{\rho_m+\rho_P}{\rho_N}+k_i$	$(w_{i_\textbf{norm}})_{\textbf{zone}}$
Central and Western	12.55	38	20057	9	2233	180	2	2235	0
						242	4	2237	0.1
						265	8	2241	0.3
						270	3	2236	0.05
						272	7	2240	0.25
						1026	11	2244	0.45
						10003	22	2255	1
						10018	5	2238	0.15
						10021	10	2243	0.4
						10046	18	2251	0.8

2.2.2. Edge weight

The number of bus routes operating between two given nodes is defined as the edge weight in the network, i.e.,

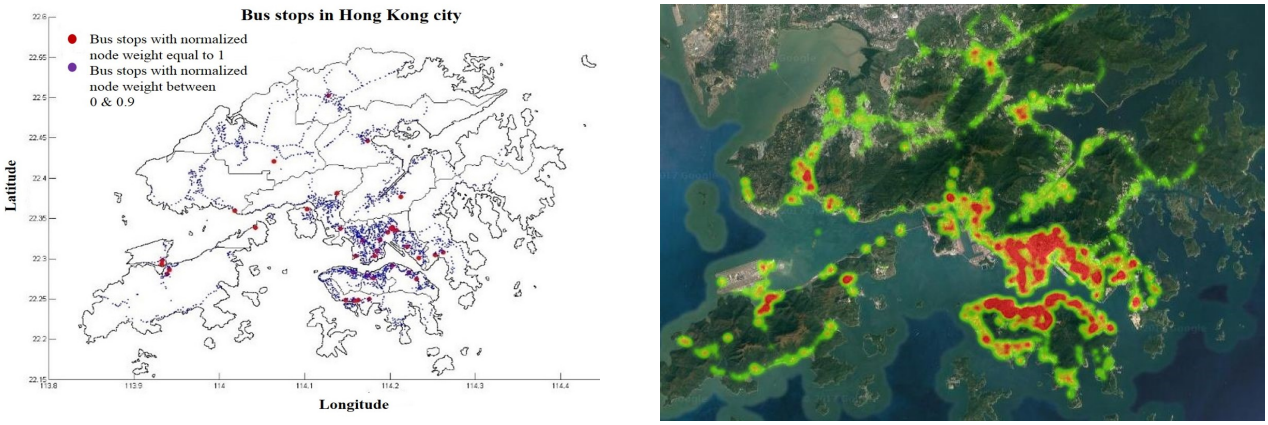
$$e_{ij} = \sum_{i \neq j} r_{ij} \tag{9}$$

where r_{ij} is the route operating between nodes i and j. The edge weight distribution for all the three cities follows an exponential distribution as shown in Fig. 5a. The long tail end of the distribution indicates that the edges with maximum weights represent the edges with most overlapped routes in the network. From the network perspective, these overlapped routes cause a few sections of the network to be over-used leading to high traffic congestions during peak travel hours. These edges also act as starting points for the traffic aggregation leading to slow moving traffic and may even cause longer waiting time for passengers to board the buses due to multiple buses waiting near the bus stops. In Fig. 5a, the distribution of $e_{ij} < 10$ is quite expected, whereas the distribution of e_{ij} between 10 to 100 describes the most overlapped routes in the network and the distribution of $e_{ij} > 100$ indicates the probabilistic occurrence of such large number of overlapped routes are mostly at central business districts (CBD) in a network. Fig. 5b shows the Bengaluru bus network structure where the maximum edge weights are observed at the city's center.



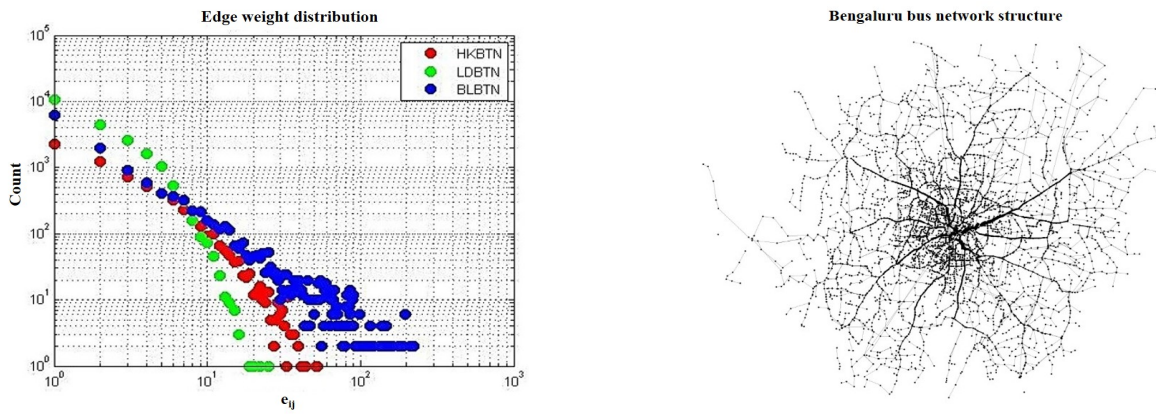
(a) (b)

Figure 3: (a) The Kowloon zone (highlighted in red color) in the Hong Kong city displaying the bus stops and POIs in the zone; (b) geographic locations of bus stops and POIs in the Hong Kong city in different zones.



(a) (b)

Figure 4: (a) Bus stop locations in the Hong Kong city with normalized node weight equal to 1 (highlighted in red); and (b) heat map showing the zones with higher demand in the Hong Kong city.



(a) (b)

Figure 5: (a) Edge weight distribution for the three cities; and (b) Bengaluru bus network with edge weights.

3. Analysis of network parameters

In this section, the weighted and directed bus network is analyzed with regular and supernode representations for the Hong Kong, London and Bengaluru networks. A detailed discussion of degree, degree distribution, clustering, mean path length, centrality measures, scale-free property, and small world behavior of the three networks is given in this section.

3.1. Degree and average degree

Degree is the most basic yet very important parameter in network analysis. The degree of a node represents the total number of edges incident on a node. For a directed network, the total degree of a node i (k_i^{total}) is equal to sum of its in-degree (k_i^{in}) and out-degree (k_i^{out})) and is given by

$$k_i^{\text{in}} = \sum_j a_{ji}, \quad k_i^{\text{out}} = \sum_j a_{ij}, \quad k_i^{\text{total}} = k_i^{\text{in}} + k_i^{\text{out}} \tag{10}$$

where a_{ij} is the adjacency matrix element corresponding to the ith node. The average degree conveys information on the average number of nearest neighbors that a node is connected to, and for a directed network it is given by

$$k_{\text{avg}} = \frac{1}{N} \sum_i k_i^{\text{total}} \tag{11}$$

Table 5: Comparison of average degree for the three networks with regular and supernode representations.

	Hong Kong	**London**	**Bengaluru**
Regular structure	2.87	1.20	2.34
Supernode structure	3.77	1.91	2.64

Table 5 shows the average degree for the three networks in regular and supernode representations. It is evident that the supernode structure increases the average connectivity of nodes in the network as compared to the regular structure. This is due to merging of nodes in the network. Moreover, this connectivity is an inherent topological behavior which can be better explained with supernode representation compared to conventional graph representation.

3.2. Degree distribution, power law and scale-free property

The degree distribution has assumed a central role in network theory following the discovery of scale-free networks [30]. The degree distribution gives the probability that a randomly chosen node in the network has degree k, i.e.,

$$p\left(k\right) = \frac{N_k}{N} \tag{12}$$

where N_k is the number of nodes with degree k and N is the total number of nodes in the network. In the network, a node with zero connectivity is called an isolated node and is often ignored. On the other hand, the node with high connectivity is called a hub. A network whose degree distribution follows a power law is called a scale-free network, i.e.,

$$p\left(k\right) \propto k^{-\gamma} \Rightarrow p\left(k\right) = Ck^{-\gamma} \tag{13}$$

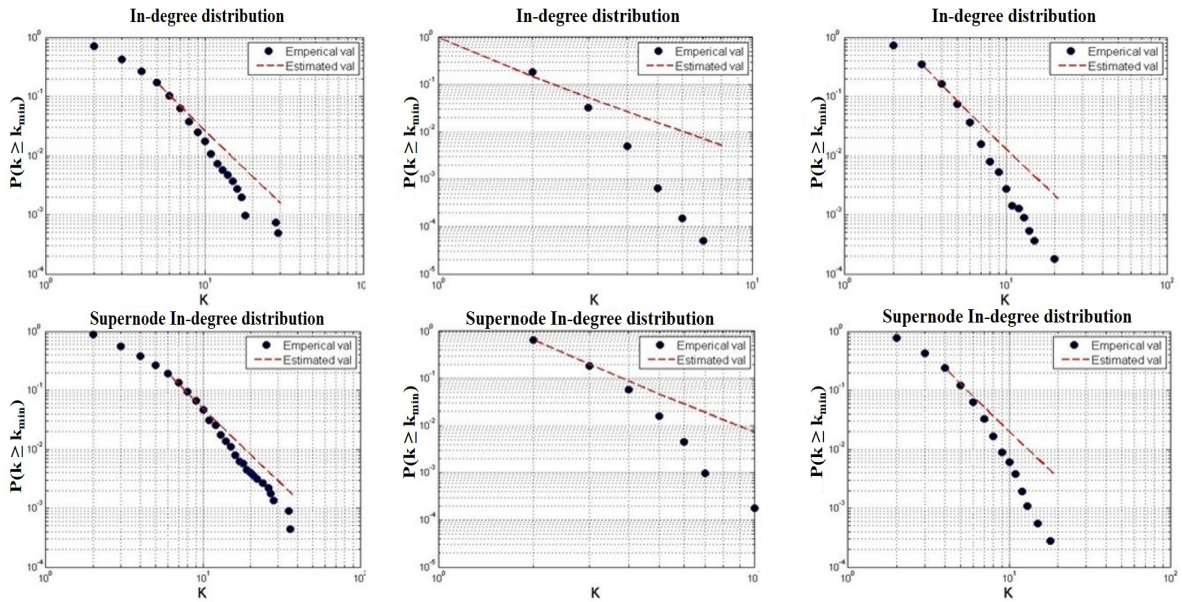
where $2 < \gamma < 3$ is the scaling parameter, k is the node degree and C is a constant. Taking log on both sides of (13), we have

$$\ln p\left(k\right) \propto -\gamma \ln k \Rightarrow \ln p\left(k\right) = -\gamma \ln k + \ln C \tag{14}$$

Thus, the power law is a straight line in log-log scale with negative slope γ. The fitting of power law distribution to the empirical data can be achieved by evaluating the value of $k_{\min}$ (the lower bound on power law) and γ (the scaling parameter) [31] through goodness-of-fit test between the empirical data distribution and the hypothesized power law distribution and compute the corresponding p-value. Typically, if the value of $p \geq 0.1$, the power law distribution is a plausible hypothesis. The graphs in Fig. 6 and Fig. 7 show the power law fit for in-degree and out-degree distributions for the three networks. It is interesting to observe that Hong Kong network fails to satisfy the power law distribution in regular network representation, but plausibly follows a power law with supernode structuring. Thus, the Hong Kong bus transport network behaves as a scale-free network with supernode structuring. However, the London and Bengaluru networks show poor fit to power law, but fit better to Poisson distribution, making the network appear to be random. Table 6 shows the values of Poisson and power law exponents for the regular and supernode structures for the three networks.

Table 6: Poisson and power law exponent values for the three networks with regular and supernode representations.

	Regular structure			**Supernode structure**		
	γ	k_{min}	λ	γ	k_{min}	λ
Hong Kong	-	-	2.87	3.5	5	-
London	-	-	1.19	-	-	1.91
Bengaluru	-	-	2.39	-	-	2.7



(a) (b) (c)

Figure 6: Power law fit for in-degree distribution for (a) Hong Kong; (b) London; and (c) Bengaluru networks in regular and supernode representations.

3.3. Clustering coefficient

Clustering coefficient, also known as the transitivity of a network, shows how well the neighbors of a node are connected to each other. For node i with degree k_i, the local clustering coefficient for a directed network is defined as [32]:

$$c_i = \frac{\sum_h \sum_j (a_{ij} + a_{ji})(a_{jh} + a_{hj})(a_{hi} + a_{ih})}{2\left[k_i(k_i - 1) - 2k_i^{\leftrightarrow}\right]} \tag{15}$$

where $k_i^{\leftrightarrow} = \sum_{i \neq j} a_{ij} a_{ji}$ and k_i is the total node degree. The global clustering is defined as:

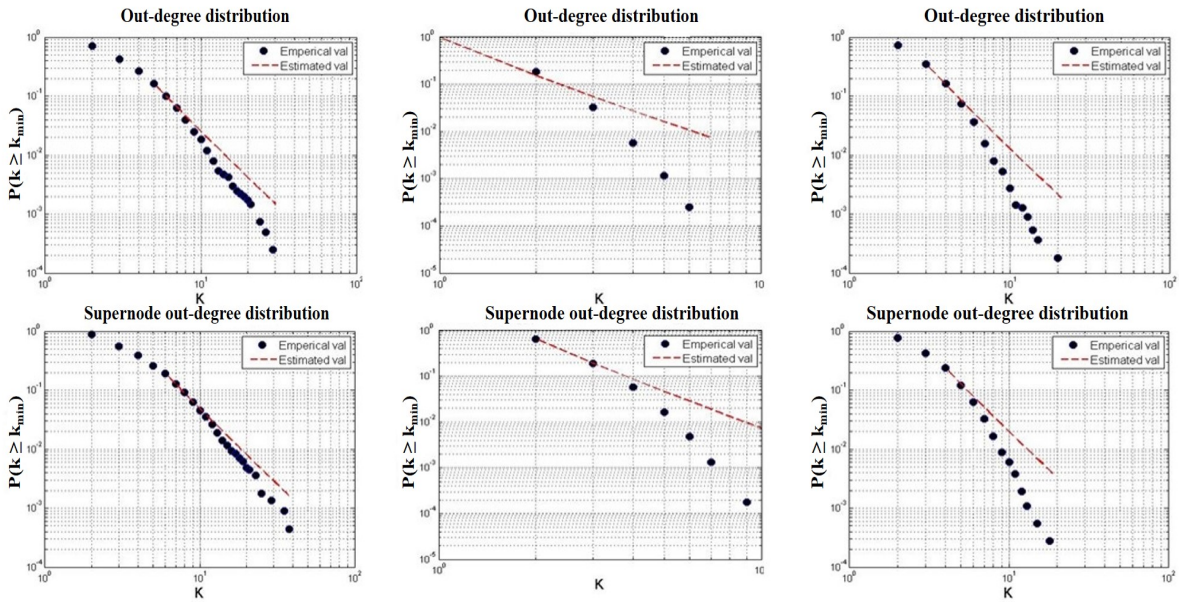
$$C = \frac{1}{N} \sum_i c_i \tag{16}$$

The degree of a node provides no information on connectivity of the nodes neighbor. The clustering coefficient measures the link density of a node's nearest neighbors by probing into the existence of triangles in the network. From (15) we can say that nodes of higher degree tend to have a lower clustering coefficient, and vice versa, since degree and clustering are inversely related. Another reason for the decrease in c_i with increasing degree is that the vertices group together tightly to form communities [33]. Fig. 8 shows the histogram for the local clustering coefficient of nodes for the three networks represented by the regular and supernode structures. It is observed that, with supernode representation, due to combining of nodes, the local connectivity significantly improves. From the network analysis viewpoint, the clustering coefficient offers better understanding of the local cohesiveness between nodes in the actual bus transport structure which provides additional useful information, like availability of alternate routes for a node at the local level during emergency. Knowing the geographical embedding of these highly-clustered nodes is an added advantage for the analysis of network behavior.

3.4. Average shortest path length

A path is a route that runs along the edges of the network and the path length represents the number of edges contained in a path between any chosen nodes i and j. The shortest path between nodes i and j is the path with the fewest number of edges between nodes i and j given by

$$L_{\text{avg}} = \frac{1}{N(N-1)} \sum_{i \neq j} d(n_i, n_j) \tag{17}$$



(a) (b) (c)

Figure 7: Power law fit for out-degree distribution for (a) Hong Kong; (b) London; and (c) Bengaluru networks in regular and supernode representations.

where $d(n_i, n_j)$ is the geodesic distance between nodes i and j, and N is the network size. Fig. 9 shows the average path length distribution for the three cities with regular and supernode representations. As observed from Fig. 9, the Hong Kong, London and Bengaluru networks follow normal distribution with a mean of 14, 175, 26, respectively with conventional graph representation, which is reduced to 8, 102, 22, respectively with supernode representation. Large values of L for the three networks at the tail of the graph is the result of considering only bus transport network in the city, and with the consideration of other transportation like metros, ferries or trams, the average path length is expected to reduce.

3.5. Small world network

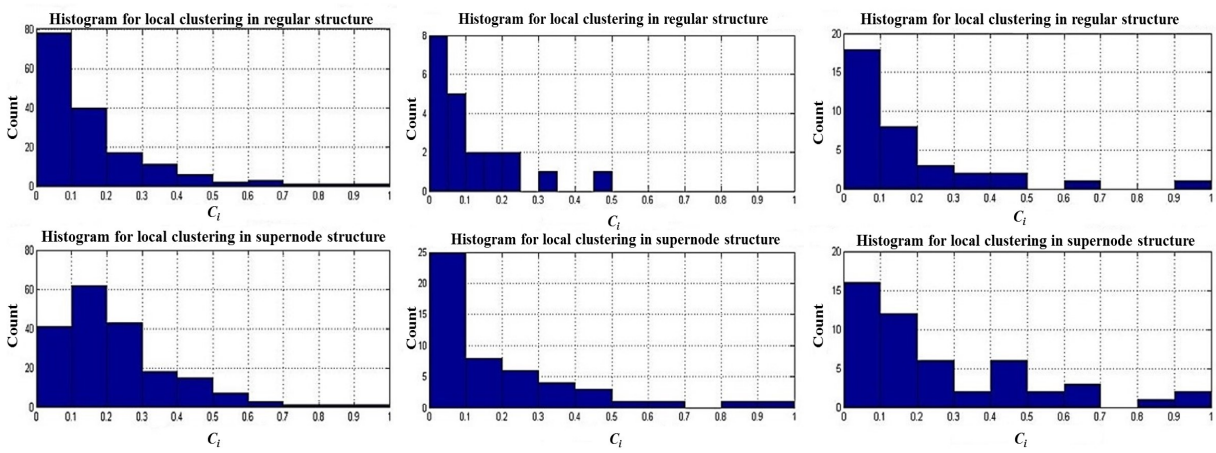
The average path length (L) and the clustering coefficient (C) are the parameters used to quantify the structural properties of a small world network. In general, a random network has poor clustering but shorter average path length, whereas a regular or lattice network has good clustering but longer path length. That is, large C is usually associated with large L, and small C is associated with small L. However, as proposed by Wattz and Strogatz [10], there exist many real-world networks with large C and small L, which are defined as small world networks. The property of small worldness is verified by calculating ω as proposed by Telesford *et al.* [34], and is given by

$$\omega(p) = \frac{L_{\text{rand}}}{L(p)} - \frac{C(p)}{C_{\text{latt}}} \quad ; 0 \leq p \leq 1 \tag{18}$$

where L_{rand} is the average path length of a random network, C_{latt} is the clustering coefficient of a lattice structure and, $L(p)$ and $C(p)$ are the average path length and clustering coefficient of the network at various rewiring probability p such that $0 \leq p \leq 1$. The graph structure is considered lattice at $p = 0$ and random at $p = 1$, and the intermediate structural behavior is analyzed for $0 < p < 1$. In the literature, a regular lattice structure is always considered as the starting point for the evaluation of small world networks. In the current work, the original bus network is considered as the starting point, and thus (18) is modified as

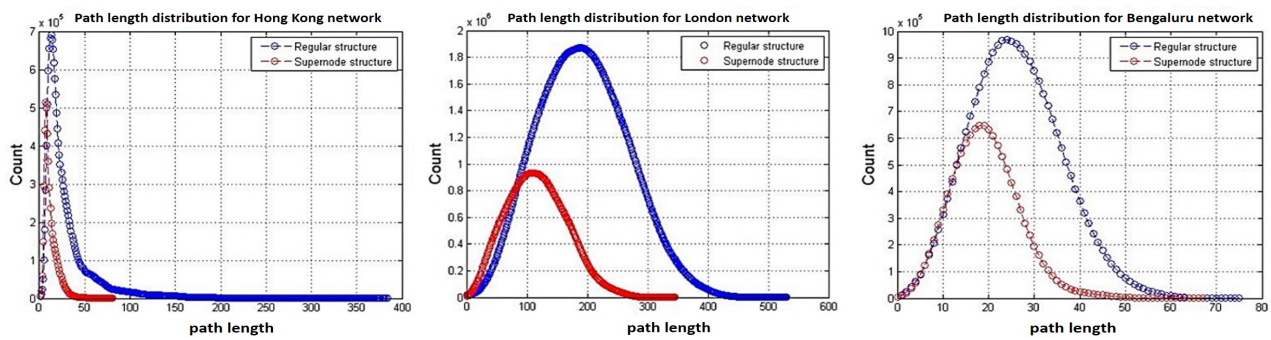
$$\omega(p) = \frac{L_{\text{rand}}}{L(p)} - \frac{C(p)}{C_{\text{existing}}} \quad ; 0 \leq p \leq 1 \tag{19}$$

The reason for considering the existing structure as a lattice structure is that a transport structure is seldom expected to be a fully connected structure. Thus, comparing the existing structure with a fully connected network is of no



(a) (b) (c)

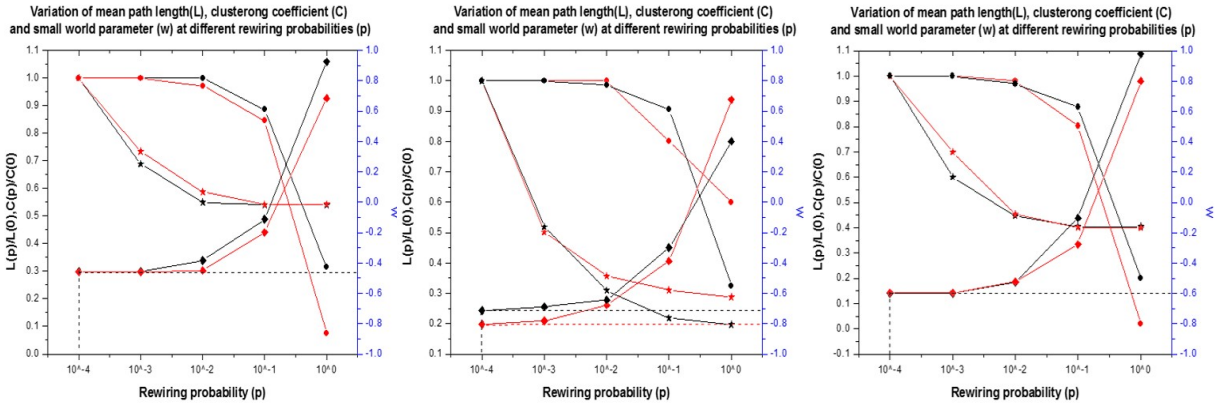
Figure 8: Histogram of local clustering coefficient for (a) Hong Kong; (b) London; and (c) Bengaluru networks.



(a) (b) (c)

Figure 9: Average path length distribution for (a) Hong Kong; (b) London; and (c) Bengaluru networks with regular and supernode representations.

practical value. From (19) the physical meaning of ω at different values of rewiring probability reflects the variation of the average path length of a network structure for a city as compared to the corresponding random network structure, since $\frac{C(p)}{C_{existing}}$ is equal to 1 in our consideration. Using the modified equation (19), the corresponding values of $L(0)$ and $C(0)$ are evaluated for the original structure. Then, the edges in the network are rewired with different values of rewiring probability p using the rewiring mechanism discussed in [35], and the corresponding values of $L(p)$ and $C(p)$ are evaluated for the three networks. Using the values of $L(p)$, $C(p)$, $L(0)$ and $C(0)$, the parameter ω is evaluated using (19) and the corresponding plots are shown in Fig. 10. The closer the value of ω to zero, the network behaves closer to six degrees of separation [34]. Table 7 shows the variation of average path length for the Bengaluru network structure at different rewiring probabilities with regular and supernode representations. Table 8 shows the values of ω for $p = 10^{-4}$ (i.e., $p \approx 0$)for all the three networks in regular and supernode representations. The values of ω for the three networks from Table 8 show that Hong Kong and Bengaluru have the potential to behave as small world networks with certain modifications to the existing routes. However, the value of ω for London remains much smaller than zero, which indicates that the network requires significant modification in the routes to behave as a small world network. However, a significant rewiring would completely change the existing network, which could be undesirable.



(a) (b) (c)

Figure 10: Small world network behavior for (a) Hong Kong; (b) London; and (c) Bengaluru networks in regular and supernode representations and the value of ω with $p = 10^{-4}$.

Table 7: Variation of L_{avg} for the Bengaluru network at different rewiring probabilities with regular and supernode representations

	p=10^{-4}	p=10^{-3}	p=10^{-2}	p=10^{-1}	p=10^{0}
Regular structure	26.07	26.06	23.59	17.57	10.5
Supernode structure	22.13	22.05	19.52	12.638	8.83

Table 8: Value of ω for the three cities in regular and supernode representations with $p = 10^{-4}$.

	Hong Kong	**London**	**Bengaluru**
Regular structure	-0.45	-0.71	-0.59
Supernode structure	-0.45	-0.80	-0.60

3.6. Centrality measures

Centrality is a measure of relative importance of a node or an edge in the network and the different definition for importance gives rise to different centrality measures. The simplest centrality is the degree centrality where the importance of a node is specified by its degree. In this section, we study a few centrality measures which are relevant to transport network analysis in L-Space modeling.

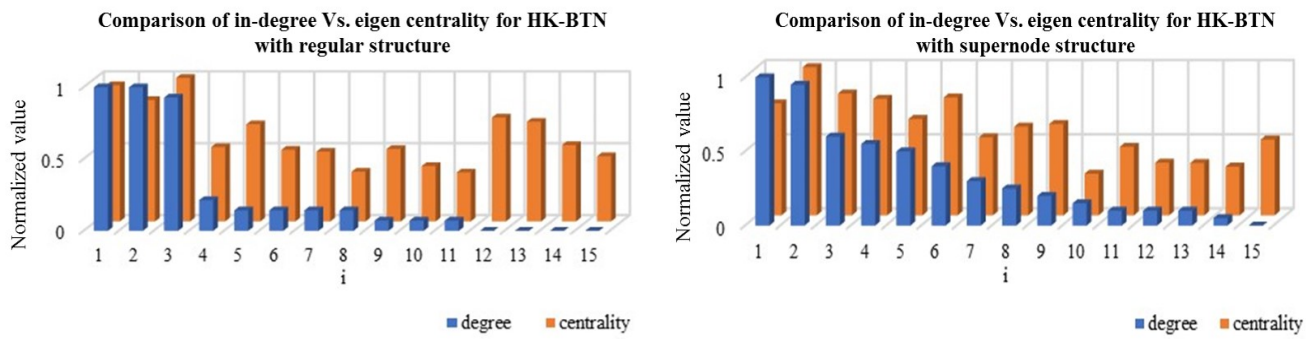
3.6.1. Eigen vector centrality

The importance of a node can be increased by having connections to other nodes which are important by themselves, i.e., a node is more important to the network not only because it has many neighbors (in-degree) but because it has important nodes as its neighbors. This is the idea behind eigen vector centrality which is mathematically given by

$$C_{n_i} = {k_i}^{-1} \sum_j a_{ji} C_{n_j} \tag{20}$$

where k_i is the largest eigen value, C_{n_i} and C_{n_j} are the eigen centrality scores for nodes i and j respectively, a_{ji} is the adjacency matrix element corresponding to node i. Fig. 11 shows a comparison of in-degree with its respective eigen vector centrality for a few nodes in the Hong Kong network. It is observed that a high in-degree node can have a low eigen centrality indicating that the node is less central as compared to some other nodes with low in-degree, but high eigen centrality indicating that the node is more central. Furthermore, the supernode structuring, along with improving the in-degree of a node, plays a significant role in identifying central nodes more accurately due to the merging of geographically closely associated nodes, hence making a few nodes more significant. That is,

supernode representation not only improves the node connectivity but also helps to identify nodes which are having connection to influential nodes. However, the problem with eigen centrality is that, if a node has only outgoing edges and no incoming edges, the centrality of this node will be zero [36]. Only a few such nodes appear in our transport networks. Thus, the eigen centrality still conveys useful information on a node's significance to the overall network.



(a) (b)

Figure 11: Comparison of in-degree and eigen vector centrality values for some nodes in the Hong Kong bus network with (a) regular; and (b) supernode representations.

3.6.2. Hub and authority centrality

A hub is a node that connects to many other nodes in the network and an authority is a node which is pointed by many hubs in the network. The Hyper Link Induced Topic Search (HITS) algorithm [15] is widely used to find hubs and authority nodes in a network. In the HITS algorithm, every node is assigned a hub weight and an authority weight. The algorithm starts by assigning an initial weight. Using this initial weight the hub and authority scores are calculated iteratively and are updated by

$$h_i \propto \sum_j (u_j a_{ij}) \quad and \quad u_i \propto \sum_j (h_j a_{ji}) \tag{21}$$

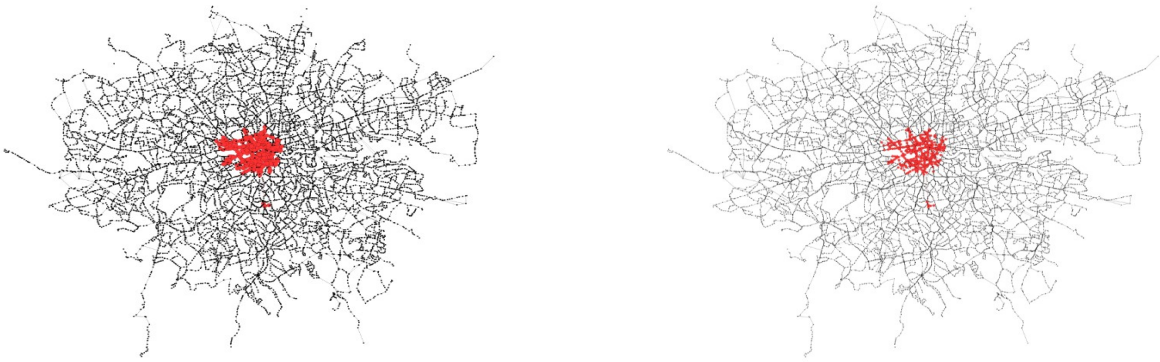
where h_i and u_i are hub and authority scores, respectively, and a_{ij} is the adjacency matrix element corresponding to node i in the network. In a BTN, hubs can be the nodes with high degree since they are connected to many nodes in the network and authorities can be nodes pointed to by these hubs, indicating that these authority nodes may cater for the highest demand in the city. Fig. 12 shows the nodes with higher order hub and authority scores for the London network, as an example. An interesting observation is that in all the three networks, the hub and authority scores for many nodes are almost the same, indicating that a hub is behaving as an authority, and vice versa.

3.6.3. Betweenness centrality

The betweenness centrality quantifies the extent to which a node acts as a bridge between any other nodes in the network and is mathematically given by

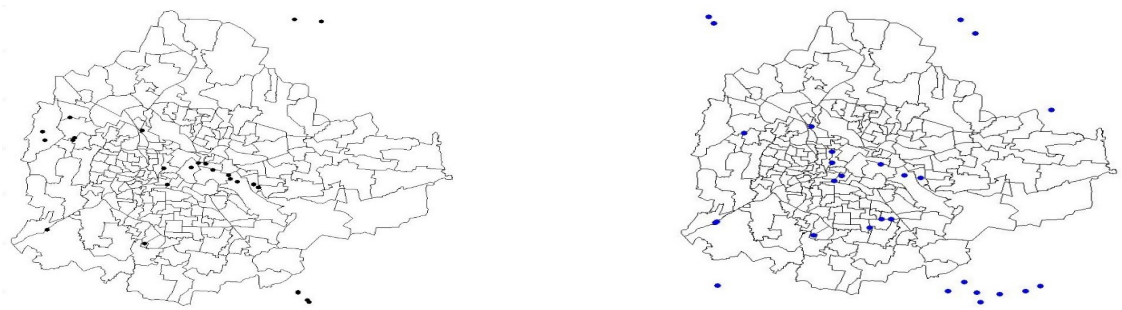
$$C_b(i) = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{22}$$

where σ_{jk} is the total number of shortest paths between nodes j and k, and $\sigma_{jk}(i)$ is the shortest paths between nodes j and k passing through node i. Fig. 13 shows the Bengaluru bus network as an example, with high betweenness centrality nodes (normalized value ≥ 0.8) in the regular and supernode representations. It can be observed that an increased number of nodes with high betweenness centrality values are better identified with the supernode representation as compared to regular representation. The physical significance of this centrality is that removing these nodes from the network since these nodes potentially control the actual routing behaviors of both passengers and buses in the network. (Note: Since the shape file (.shp) shows only Bengaluru urban zone and not the complete urban and rural zones, some nodes lie outside the network).



(a) (b)

Figure 12: The London bus network showing the nodes with higher (a) hub score; and (b) authority score.



(a) (b)

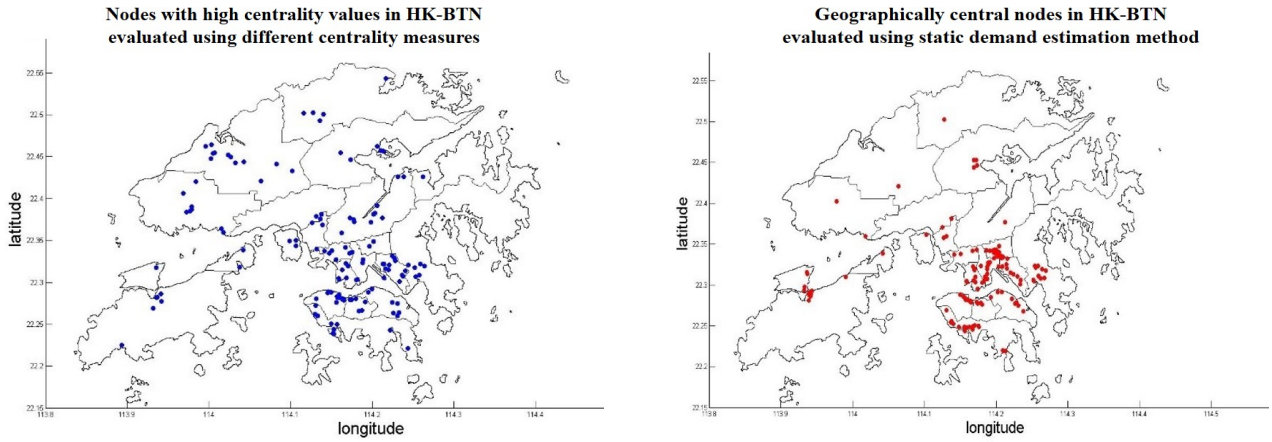
Figure 13: Bus stops in Bengaluru network with high betweenness centrality (normalized value ≥ 0.8) in (a) regular; and (b) supernode representations.

3.6.4. Figure of merit

In this section, the centrality of a node is compared with its geographical centrality to verify the importance of modeling node weight in the transport analysis. For example, Fig. 14 shows the comparison of highly central nodes considered in Section 3.6 versus the nodes which are considered geographically more significant as evaluated using (7). From Fig. 14b it is observed that the geographical significance of the node evaluated at the local level using (7) yields almost the same result as estimating different centrality measures at the global level, as shown in Fig. 14a. That is, a set of 130 nodes were identified in the Hong Kong supernode structure constituting the higher order values (normalized value > 0.8) for degree centrality, hub centrality, betweeneness and eigenvector centrality, which are considered highly central nodes in the network. On the other hand, a set of 130 nodes with the high node weight values (normalized value > 0.8) evaluated using (7) were identified. The comparison of node centrality and geographical significance of the chosen 130 nodes showed almost 80 nodes in common indicating that the proposed demand estimation method using (7) is a better way to assign node weights to signify the geographical importance of a node. The correlation between topologically central nodes and the geographically central nodes reveals the interesting fact that the proposed static demand estimation method for assigning node weights aids in better identifying the geographically significant nodes in the network

4. Topological efficiency

In Section 3, we analyzed different structural properties of the three networks to gain better insight into their topological behavior. In this section, we first analyze the topological efficiency of the network in terms of distance, which reflects the inherent contribution of the network to efficiency. We then propose an efficiency modeling in



(a) (b)

Figure 14: Hong Kong bus transport network in supernode representation with highly central nodes (a) evaluated using different centrality measures; and (b) evaluated using static demand estimation method.

terms of time, rather than distance, which is more practical for assessing the efficiency of end-to-end travel in a transport network. As per network theory, topological efficiency of a network is given by

$$\eta_G = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{23}$$

where d_{ij} is the shortest distance path between nodes i and j, and N is the network size. Table 9 shows the comparison of the topological efficiency of the three networks in regular and supernode structures. From Table 9, it also evident that the Hong Kong network is topologically more efficient than the London and Bengaluru networks. Also, the efficiency of the three networks is higher in supernode structure as compared to the regular structure. In fact, we can say that the actual efficiency of the network is better explained with supernode structuring. Although no physical changes are made to the actual network, a slight restructuring in network representation helps to gain better insight into the actual behavior of the network.

Table 9: Comparison of topological efficiency between three structures in regular and supernode representations.

η_G	**Hong Kong**	**London**	**Bengaluru**
Regular structure	0.099	0.015	0.038
Supernode structure	0.115	0.023	0.057

Equation (23) gives the topological efficiency of the network, conveying information on whether the network offers shorter distance of travel between any two given nodes on average. However, when a passenger prefers to travel between two nodes, the end-to-end travel delay between the nodes is taken into consideration rather than the distance. Of course, other parameters like minimum transfer, cost effectiveness and convenience are also considered. In the current work, (23) is remodeled in terms of delay ($\eta_{G,t}$) rather than distance to measure the topological efficiency and is given by

$$\eta_{G,t} = \frac{1}{N(N-1)} \sum_{\substack{i=1\ldots n-1 \\ j=i+1\ldots n}} \frac{d_{ij}}{v_{ij}} \tag{24}$$

In (23), d_{ij} is the total number of hops between nodes i and j, but in (24) d_{ij} measures the actual geographic distance of every hop along the shortest distance between nodes i and j. Also, v_{ij} is the average velocity of every hop along the shortest path between nodes i and j. The term $\frac{d_{ij}}{v_{ij}}$ gives the end-to-end travel delay time between nodes i and j. Equation (24) is validated by carrying out a simple simulation in SUMO (Simulation of Urban Mobility) for the Hong Kong network.

4.1. SUMO simulations

SUMO is a microscopic multi-modal traffic simulator which allows the behavior of each vehicle to be explicitly controlled and monitored [37]. In this section, a simple simulation is conducted using SUMO to evaluate the topological efficiency in terms of the end-to-end delay using (24). Details of the simulation set up are discussed below:

Step 1. Building a road network topology:
To build a road network in SUMO, we can either import the network from Openstreetmap [38], or we can build the network manually. In the current work, we have built the network topology manually. The bus stops are treated as nodes and the road segment (including the number of lanes per road segment) connecting two bus stops is treated as an edge. The geographical information of the bus stops, the lanes per road segment, the pedestrian crossing, and traffic light signal, are extracted from the Openstreetmap. All the traffic lights are generated with a default cycle of 90 s (40 s green, 5 s amber, 40 s red followed by 5 s amber to switch to the next cycle of Green). The simulation in this paper considers the road topology as described above for one specific bus route (Route No. 1) between the chosen source nodes S (CHUK YUEN ESTATE) and the destination node D (STAR FERRY). The end nodes S and D, and the specific route between them is chosen for the simulation since the route passes through different zones of the city with significant POIs, which helps to verify the nodal weight analysis as discussed in Section 2.2.1.

Step 2. Setting up the routes for both buses and other vehicles in the network:
The route details and the frequency for the bus route are configured according to the official timetable in [39]. We use activitygen to generate the traffic other than buses [40], which is a tool in the SUMO simulator for generating traffic in a network based on the activity in a zone. It makes use of the activity-based traffic model to generate such traffic and is also known as activity based demand generator. An activity can be regarded as a trip going to or from an office, school, or free time travel. By providing the input data on the number and locations of POIs, number of people living in the zone, number of people traveling from other zones to the chosen zone for their work, and the working hours in a day, activitygen generates activities happening in the zone. Using another tool called Duarouter [41], every activity in the zone is assigned by a route based on selecting the shortest path between the source and the destination. The destination could be a POI that the passenger wants to reach and the source could be a location within the zone or outside the zone where the passenger starts the journey. If the passenger starts the journey beyond the chosen zone, a specific location through which he enters the given zone can be explicitly controlled in SUMO. The routes generated by Duarouter are for vehicles other than buses in the network. These vehicles enter the network from a source location which is manually configured in SUMO (in the current simulation, the source location is configured as the location around the starting point of the chosen bus route, i.e., S) and leave the network from a specific location (manually configured to be around the bus stop location D) and remain in the network until the specified time configured in activitygen. Fig. 15a shows a snapshot of the simulation setup described above, indicating the bus stops, traffic light signals, pedestrian walkways and POIs. Fig. 15b shows a snapshot of vehicles moving in the network at the time instance of 29400 s (8:10 AM).

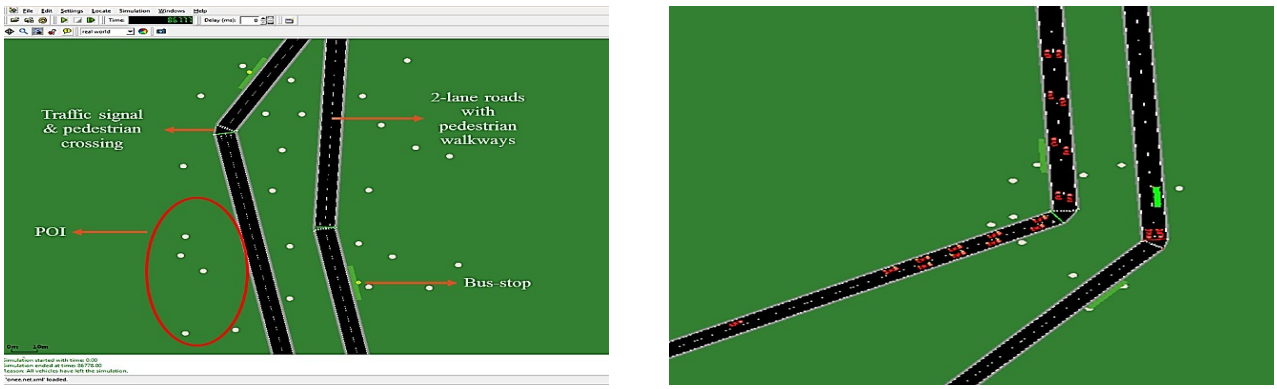
Step 3. Simulate the scenario considering the road network as described in step 1 and the route set up as described in step 2 for the morning peak hour (8:00 to 9:00 am).

Step 4. Extract the results from the SUMO output files:
Since SUMO is a microscopic simulator, it can log the results of individual vehicles in a trace file with a sampling period of one second. We can also extract the information on the geographical position of a vehicle, the route information, the velocity of a vehicle, etc. Using the trace file generated at the end of the simulation in step 3, the time mean speed of the vehicles (buses) are extracted for every road segment for an hour.

Step 5. Calculate the end-to-end travel:
The end-to-end travel delay is calculated between the two chosen nodes S and D using (24) where, the geographic distance between the bus stops is calculated using the Haversine formula [27], and the time mean speed is calculated as described in step 4. The final result of our simulation set up is compared with the Google map's expected travel time, the Hong Kong's official e-Transport application and the empirical result from the real-world KMB (Kowloon Motor Bus Co.) dataset as tabulated in Table 10.

Table 10: Comparison of end-to-end travel delay between chosen nodes S and D during morning peak hour.

Parameter	HKeTransport	Google Maps	Simulation result	Emperical result
end-to-end delay (min)	65	57	56.5	52



(a) (b)

Figure 15: (a) Snapshot of the SUMO simulator; and (b) vehicles moving in the network.

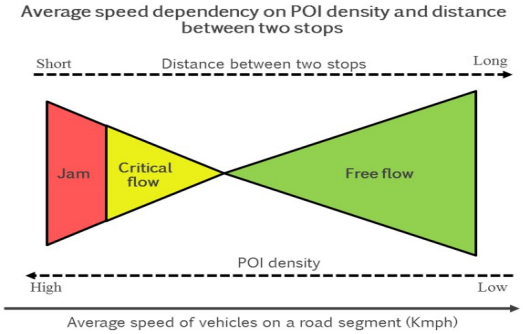
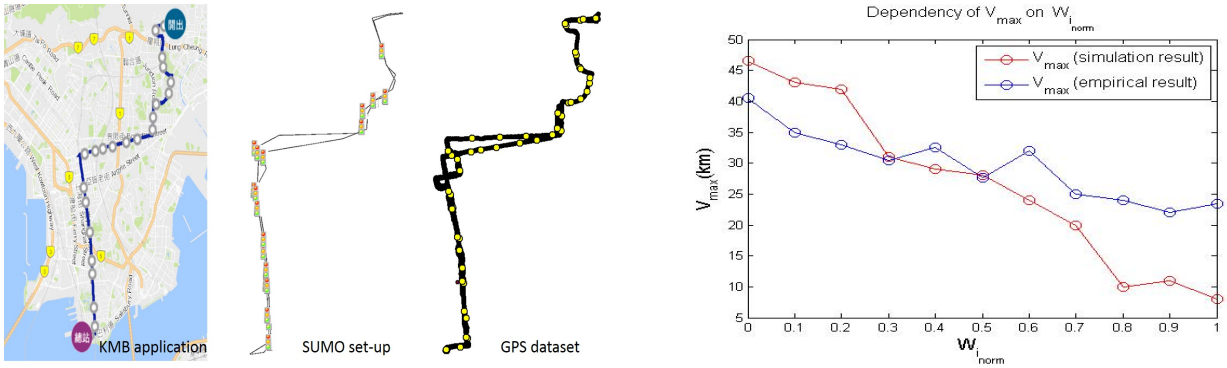


Figure 16: Generalized flow prediction for dependency of vehicle speed on POI density and distance between the stops.



(a) (b)

Figure 17: (a) The comparison of route details for the chosen bus route no. 1 as shown in the KMB application, SUMO simulation set-up and the GPS locations obtained from empirical dataset. (The bus stops in the SUMO simulation are not highlighted, instead, the locations of the traffic light signals configured in the simulation are highlighted); and (b) comparison of the simulation and empirical result for the dependency of vehicular speed on node weight.

Another important thing to notice in the simulation is the dependency of the vehicular speed along a road segment with respect to the node weight (POI density and the node occupying probability as discussed in Section 2.2.1). In our simulation, we observe that, as the number of POIs around a particular bus stop increases, the maximum speed attained by the vehicles in the particular road segment decreases, considering the peak hour simulation. The situation goes worst when the distance between the bus stops decreases and the node weight increases, i.e., as shown in Fig. 16, the speed of the traffic changes from free flow to critical flow, and finally to a jammed state as the node weight (POI density and node occupying probability) increases, and the distance between the bus stops decreases. To validate our simulation results, a real-world data provided from the KMB (Kowloon Motor Bus Co.) operator (from Hong Kong) is employed. The data is collected for a period of one week during October 2017 for the morning (8:00-11:00 AM) and evening peak hours (4:00-7:00PM). The data x (latitude), y (longitude), t (time), v (velocity at a given time) is sampled with a sampling rate of one second along the chosen route (Route no. 1). Fig. 17a shows the GPS (Global Positioning System) locations for the chosen bus route no. 1. Using the data, the maximum speed achieved along a road segment is extracted for both morning and evening peak hours considering an average of 30 trips along the chosen route (a road segment is considered as a segment between successive bus stops, i.e., it can be a collection of smaller road segment separated by road junctions or traffic light signals). Fig. 17b shows the dependency of maximum speed attained along a road segment ($V_{\max}$) on the normalized node weight (w_{i_norm}). The comparison of simulation and empirical data show that as node weight increases, the $V_{\max}$ decreases. This condition might not be true always, unless we consider the distance between the successive nodes. Fig. 18 plots distance between any two successive stops (d_{ij}), the normalized node weight (W_{i_norm}) and the maximum speed achieved along a road segment ($V_{\max}$) for the morning and evening peak hours. We can see from the figure that, the normalized node weight is the node occupying probability, i.e., the number of people accessing a particular stop as per the real-world data from KMB is considered to be the demand serviced by a node. From Fig. 18, we can notice that the bus stops no. 9, 10 and 11 (S9, S10, S11) have relatively higher node weight, however, since the stops S9 and S10 are geographically closer, the $V_{\max}$ along the road segment is smaller as compared to the road segment between the stops S10 and S11. A similar scenario is observed between the stops S17, S18 and S19. Thus, from our findings we can infer that, with increasing node weight (demand) and geographically closer bus stops, the maximum speed achieved along the road segment reduces significantly. It is acceptable that the operators deploy more stops to meet the greater demand, but it should also be considered that closer bus stops would eventually lead to a state of traffic congestions. Hence, the route planning and the stop deployment needs judicious design to overcome the congested travels. Additionally, as shown in Fig. 19, considering the practical demand serviced by the nodes along the chosen route, it is observed that though certain stops have lower node weight (demand), their weighted node degree (the number of bus routes servicing the stop) is significantly high. Such a scenario is observed between stop S19 to S25. On the contrary, there are stops with higher demand and are serviced by less number of routes, as observed between stops S9 to S11. Thus, the node weight analysis as discussed in Section 2.2.1 aids the bus operators to not only plan their routes in a better way considering the node weight, but also assists in choosing the location of stops along a route. Though the simple simulation carried out in Section 4.1 is primarily to get an idea on the analysis of topological efficiency in terms of end-to-end travel delay, the simulation gives us the basic idea on the dependency of $V_{\max}$ on W_{i_norm} and d_{ij} which is validated with real-world empirical data.

5. Conclusion and future work

The analysis of the topological properties and structural behavior of Hong Kong, London and Bengaluru bus transport networks in regular and supernode representations revealed that the Hong Kong network is topologically more efficient as compared to the London and Bengaluru networks. The degree distribution of the Hong Kong network follows a power law, whereas an exponential distribution is observed for London and Bengaluru networks, indicating that only Hong Kong network is a scale-free network. An improved level of clustering has been noticed in all the three networks due to the merging of nodes. This property reduces the average shortest path required to travel between two nodes. The Hong Kong and Bengaluru networks with their average path length close to six degrees of separation under supernode structuring need minor route modifications to behave as small world networks. However, the London network is extremely sparse. The geographically significant nodes identified with the aid of the demand estimation model indicate the importance of the nodes to the network in local zones. Interestingly, many of these nodes are found to be central nodes in the network from a graph theory perspective. Thus, unlike previous approaches, which used different centrality measures to evaluate the significance of nodes that gives an

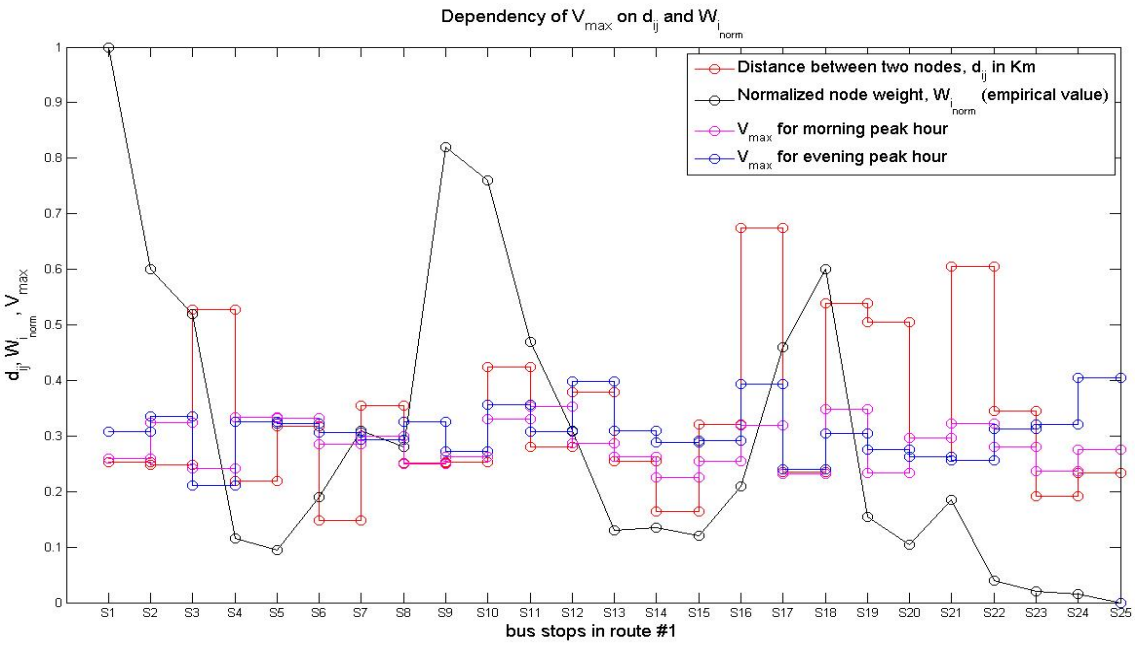


Figure 18: Dependency of the maximum speed achieved along a road segment ($V_{\max}$) on distance between any two successive stops (d_{ij}) and the normalized node weight (W_{i_norm}) for the morning and evening peak hours along the bus route no. 1. The values of $V_{\max}$ are scaled by a factor of 100 to ensure the data fits into the range of [0:1] along the y-axis.

idea on the static node behavior, in this work, we assign node weight based on the POIs and population density to capture the dynamic behavior of nodes in the network, i.e., by understanding the variations in population density and POI statistics around a node, a better estimate on its dynamic behavior can be obtained. By understanding such dynamic behavior of nodes, bus operators can better plan their routes to avoid a congested travel. Finally, the end-to-end delay model offers a better definition to measure the actual structural efficiency in terms of time as compared to the distance term employed earlier. The effect of dependency of points-of-interest density and the distance between two bus stops based on the travel speed shows that the location of the points-of-interest around a bus stop has a major influence not only on its demand estimation but also on the end-to-end travel delay or topological efficiency. As part of future work, the static demand estimation model is to be verified for the London and Bengaluru networks. The node occupying index $\frac{P_j}{N_j}$ considered in the current work is a static and uniform value across different zones for the time being, which can be made more practical by training with real-world data on node usage and points-of-interest using deep learning methods. Lastly, the real-time traffic speed can be used as the edge weight to analyze the actual network behavior at different time instants throughout the day.

Acknowledgement

This work is partially supported by the National Natural Science Foundation of China (Project No. 61401384); and The Hong Kong Polytechnic University (Projects 4-BCCH, 4-ZZCZ).

References

[1] A.-L. Barabási, Network science, Cambridge university press, 2016.

[2] S. Feng, B. Hu, C. Nie, X. Shen, Empirical study on a directed and weighted bus transport network in china, Physica A: Statistical Mechanics and its Applications 441 (2016) 85–92.

[3] X. Zhang, G. Chen, Y. Han, M. Gao, Modeling and analysis of bus weighted complex network in qingdao city based on dynamic travel time, Multimedia Tools and Applications 75 (24) (2016) 17553–17572.

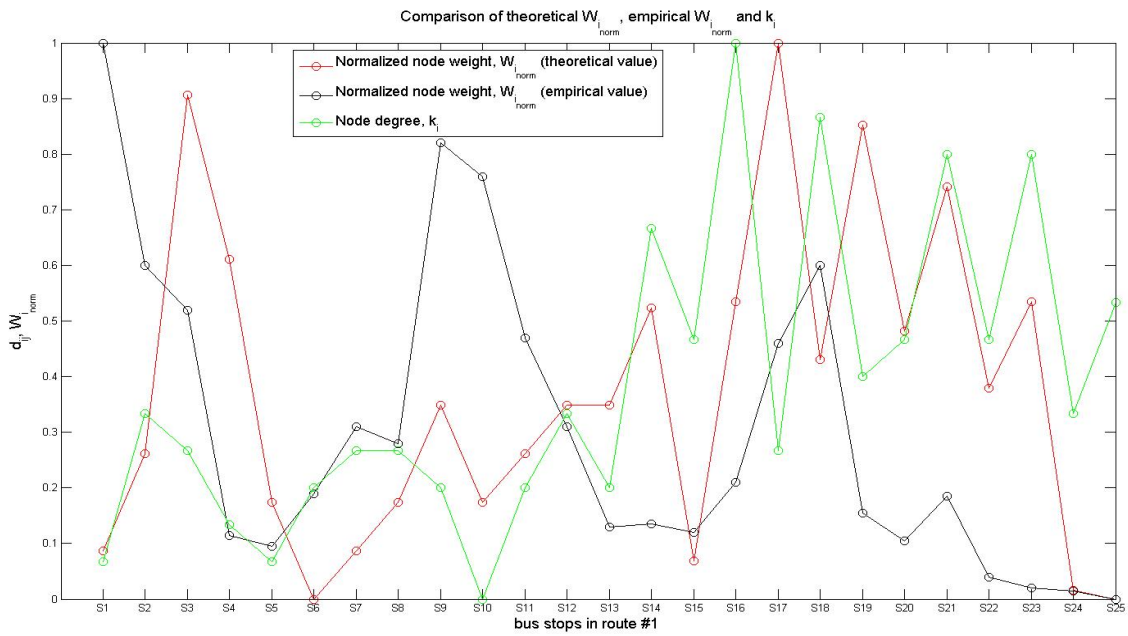


Figure 19: Comparison of theoretical node weight, empirical node weight and their corresponding node degree.

[4] Hong Kong Historical Dataset, `https://data.gov.hk/en-data/category/transport?organization=hk-td/`, [Online; accessed 1-Feb-2016].

[5] Transport for London Dataset, `http://content.tfl.gov.uk/london-bus-network-statistics.pdf/`, [Online; accessed 10-Jan-2017].

[6] C. G. Anand, Presentation on bus-based public transport in bangalore bmtc experience (2013) 1–63.

[7] OpenCity Urban Data Portae, `http://opencity.in/topic/transportation/`, [Online; accessed 16-April-2017].

[8] P. Erdos, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci 5 (1) (1960) 17–60.

[9] B. A.-L. Barabási, E. Bonabeau, Scale-free, Scientific American 288 (5) (2003) 50–59.

[10] D. J. Watts, S. H. Strogatz, Collective dynamics of'small-world'networks, nature 393 (6684) (1998) 440.

[11] R. Burt, The social structure of competition. n. nohria, r. eccles, eds, Networks and Organizations: Structure, Form, and Action (1992) 57–91.

[12] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43.

[13] P. Bonacich, Power and centrality: A family of measures, American journal of sociology 92 (5) (1987) 1170–1182.

[14] S. Brin, L. Page, The pagerank citation ranking: bringing order to the web.

[15] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM (JACM) 46 (5) (1999) 604–632.

[16] V. Latora, M. Marchiori, Is the boston subway a small-world network?, Physica A: Statistical Mechanics and its Applications 314 (1) (2002) 109–113.

[17] J. Sienkiewicz, J. A. Hołyst, Statistical analysis of 22 public transport networks in poland, Physical Review E 72 (4) (2005) 046127.

[18] Y.-Z. Chen, N. Li, D.-R. He, A study on some urban bus transport networks, Physica A: Statistical Mechanics and its Applications 376 (2007) 747–754.

[19] X. Xu, J. Hu, F. Liu, L. Liu, Scaling and correlations in three bus-transport networks of china, Physica A: Statistical Mechanics and its Applications 374 (1) (2007) 441–448.

[20] C. Von Ferber, T. Holovatch, Y. Holovatch, V. Palchykov, Public transport networks: empirical analysis and modeling, The European Physical Journal B-Condensed Matter and Complex Systems 68 (2) (2009) 261–275.

[21] A. Chatterjee, M. Manohar, G. Ramadurai, Statistical analysis of bus networks in india, PloS one 11 (12) (2016) e0168478.

[22] H. Zhang, P. Zhao, J. Gao, X.-m. Yao, The analysis of the properties of bus network topology in beijing basing on complex networks, Mathematical Problems in Engineering 2013.

[23] H. Soh, S. Lim, T. Zhang, X. Fu, G. K. K. Lee, T. G. G. Hung, P. Di, S. Prakasam, L. Wong, Weighted complex network analysis of travel routes on the singapore public transportation system, Physica A: Statistical Mechanics and its Applications 389 (24) (2010) 5852–5863.

[24] Wikipedia, Arcgis — wikipedia, the free encyclopedia, [Online; accessed 31-July-2017] (2017). URL `https://en.wikipedia.org/w/index.php?title=ArcGIS&oldid=789615115`

[25] Wikipedia, Google earth — wikipedia, the free encyclopedia, [Online; accessed 31-July-2017] (2017). URL `https://en.wikipedia.org/w/index.php?title=Google_Earth&oldid=792723706`

[26] Wikipedia, Gephi — wikipedia, the free encyclopedia, [Online; accessed 31-July-2017] (2017). URL `https://en.wikipedia.org/w/index.php?title=Gephi&oldid=787601739`

[27] Haversine formula- The great circle distance between two points on a sphere given their latitudes and longitudes, `https://en.wikipedia.org/w/index.php?title=Haversine_formula&oldid=779367200/`, [Online; accessed 10-Jul-2017].

[28] C. F. Daganzo, Structure of competitive transit networks, Transportation Research Part B: Methodological 44 (4) (2010) 434–446.

[29] GeoInfoMap - A spatial information service by Hong Kong Special Administrative Region , `http://www1.map.gov.hk/gih3/view/index.jsp/`, [Online; accessed 1-Feb-2016].

[30] A.-L. Barabási, Scale-free networks: a decade and beyond, science 325 (5939) (2009) 412–413.

[31] A. Clauset, C. R. Shalizi, M. E. Newman, Power-law distributions in empirical data, SIAM review 51 (4) (2009) 661–703.

[32] G. Fagiolo, Clustering in complex directed networks, Physical Review E 76 (2) (2007) 026107.

[33] M. E. Newman, Detecting community structure in networks, The European Physical Journal B-Condensed Matter and Complex Systems 38 (2) (2004) 321–330.

[34] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, P. J. Laurienti, The ubiquity of small-world networks, Brain connectivity 1 (5) (2011) 367–375.

[35] S. M. e. a. Ruowen Liu, Porter Beus, Analysis of watts-strogatz networks (2015) 1–6.

[36] M. Newman, Networks: an introduction, Oxford university press, 2010.

[37] M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, Sumo–simulation of urban mobility: an overview, in: Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation, ThinkMind, 2011.

[38] Openstreetmap , `https://www.openstreetmap.org/#map=5/51.500/-0.100/`, [Online; accessed 13-Nov-2015].

[39] Kowloon Motor Bus (KMB) route search, http://search.kmb.hk/KMBWebSite/index.aspx?lang=en/, [Online; accessed 21-Jun-2017].

[40] Activitygen , http://sumo.dlr.de/wiki/Demand/Activity-based_Demand_Generation/, [Online; accessed 13-Nov-2015].

[41] Activitygen , http://www.sumo.dlr.de/userdoc/DUAROUTER.html/, [Online; accessed 13-Nov-2015].