The following publication T. Shanmukhappa, I. Wang-Hei Ho, C. K. Tse and K. K. Leung, "Recent Development in Public Transport Network Analysis From the Complex Network Perspective," in IEEE Circuits and Systems Magazine, vol. 19, no. 4, pp. 39-65, Fourthquarter 2019 is available at https://dx.doi.org/10.1109/MCAS.2019.2945211.

Recent Development in Public Transport Network Analysis from the Complex Network Perspective

Tanuja Shanmukhappa, *Student Member, IEEE*, Ivan Wang-Hei Ho, *Senior Member, IEEE*, Chi K. Tse, *Fellow, IEEE*, and Kin K. Leung, *Fellow, IEEE*

***Abstract*—A graph, comprising a set of nodes connected by edges, is one of the simplest yet remarkably useful mathematical structures for the analysis of real-world complex systems. Network theory, being an application-based extension of graph theory, has been applied to a wide variety of real-world systems involving complex interconnection of subsystems. The application of network theory has permitted in-depth understanding of connectivity, topologies, and operations of many practical networked systems as well as the roles that various parameters play in determining the performance of such systems. In the field of transportation networks, however, the use of graph theory has been relatively much less explored, and this motivates us to bring together the recent development in the field of public transport analysis from a graph theoretic perspective. In this paper, we focus on ground transportation, and in particular the *bus transport network* (BTN) and *metro transport network* (MTN), since the two types of networks are widely used by the public and their performances have significant impact to people's life. In the course of our analysis, various network parameters are introduced to probe into the impact of topologies and their relative merits and demerits in transportation. The various local and global properties evaluated as part of the topological analysis provide a common platform to comprehend and decipher the inherent network features that are partly encoded in their topological properties. Overall, this paper gives a detailed exposition of recent development in the use of graph theory in public transport network analysis, and summarizes the key results that offer important insights for government agencies and public transport system operators to plan, design, and optimize future public transport networks in order to achieve more efficient and robust services.**

Keywords: Public transport network analysis, bus transport networks, metro transport networks, network science, graph theory, survey.

I. Introduction

Public transportation systems form a vital part of our infrastructure that permits massive flow of commuters within a city and between cities. In order to meet the rising standards of living of the society, transportation networks have to keep abreast of the need of commuters with respect to the ever increasing demand of reducing the traveling time and extending the area covered. At the same time, transportation networks are facing series of challenges, including satisfying the ever increasing passenger volume, achieving long-term sustainability, and improving the quality of service. Such challenges are encountered at various levels of operation, ranging from infrastructure deployment to optimal route planning, and the problems are addressed from different angles depending on the discipline of study such as urban planning, regional science, geography, engineering, etc. The literature abounds with diverse methodologies adopted in various disciplines to represent, perceive and analyze the complex dynamics of public transport systems, among which, Geographic Information System (GIS), graph theory, mathematical programming, and agent-based modeling are most commonly adopted [1].

Motivated by the notable contributions of network theory [2], application of graph theoretic concepts in the analysis of public transport networks (PTN) has attracted significant attention, and today, it is one of the most widely employed approaches to understand the nature of connectivity in PTNs. The representation of a PTN as a complex network, together with the adoption of some concepts from statistical physics, offers remarkable advantages in the modeling and analysis of nonlinear and dynamic PTN structures. Specifically, the analysis of PTNs using network theory permits the use of a common platform on which to comprehend and decipher the inherent network features that are encoded in the topological properties. Moreover, to apply the concepts of complex networks, one should understand the language of graph theory, as a prerequisite, where a network is typically represented as a graph consisting of a set of *nodes* interconnected by a set of *edges*.

Graph theory and network theory, despite being rooted historically in mathematics, has found applications in statistical physics, biology, social sciences, finance, and engineering. One of the oldest instances of using the notion of graph theory to analyze a real-world problem dates back to the 17th Century when Leonhard Euler used the concept of nodes and edges to solve the problem of seven bridges of Königsberg, a notable problem in the history of mathematics [3]. However, notable usage of graph theory was found by Gustav Kirchhoff who employed nodes and edges to

T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse are with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong.

K. K. Leung is with the Department of Electrical and Electronic Engineering, and Computing, Imperial College London, United Kingdom.

calculate voltages and currents in electric circuits, nowadays widely known as Kirchhoff's laws [4]. Subsequently, many real-world networks were analyzed using graph theory with significant contributions from the fields of social networks (world wide web) and biological networks, and later from other fields including friendship networks, relationship in social media, food web, metabolism, professional ties, author and co-author relations, citation networks, computer virus flow, network router analysis, chemical reactions, neural networks, transportation networks, etc. From the literature, it was evident that modeling various large real-world network structures as graphs, and analyzing their behavior from a network perspective, facilitated better understanding of both the global and local properties of the network. Thus, this domain of study has attracted a humongous amount of research interest in the past two decades [5], [6], [7]. Although a lot of real-world complex systems have been analyzed using graph theory, little attention has been paid to the field of transportation networks which is an active research area among researchers in transportation and logistics.

Although a public transport network can either be uni-modal or multi-modal, we focus on two major types of public transportation, namely, the bus transport network (BTN) and metro transport network (MTN), since we believe that these two types of networks are most widely used by the public to meet their daily commuting needs. In this article, the recent contributions and the concepts employed in the topological analysis of public transport networks are discussed. Our focus is on the understanding of various network parameters and approaches employed to analyze the topology of a PTN [5], [6], [7]. Moreover, a brief discussion on the fundamental graph theoretic concepts will be made whenever necessary.

The remainder of the paper is structured as follows. Section II introduces a few preliminary steps to be followed to construct a real-world network topology from given datasets, i.e., collecting the real-world datasets from various online sources, and data mining to extract useful information from both computational and visualization perspectives. Section III presents various spaces of graph representation for studying the topological representation of PTNs. Section IV discusses in detail the contributions of previous works in terms of the use of appropriate network parameters that aid PTN analysis. Section V focuses on some distinctive contributions accomplished in PTN analysis which might pave the way for future research or some food for thought. Finally, in Section VI, a few important conclusions are drawn, and the possible scope of future work is discussed.

II. Data Collection, Mining and Visualization

Although significant research interest in the field of network science theory has been cultivated for several decades, applying the established concepts to real-world data has been practiced only recently as a consequence of the availability of real-world datasets and the high-end tools for processing such huge datasets. With the aid of real-world datasets, a network topology which closely mimics the real-world structure is generated using the concepts of graph theory. Building a network topology forms the fundamental and important aspect in a PTN analysis since the course of the defined topology significantly influences the understanding of both the local and global aspects of a network. A list of online sources and relevant datasets are given in Appendix A. The extracted datasets include information on

(i) List of *stops/stations* along with their id's, names, latitude/northing and longitude/easting data.
(ii) List of *routes/sequence-of-stops* along with their stop sequence id's and names for the inbound, outbound, and round-trip routes.

where a *stop* or *station* is a designated place allocated to pick up or drop off passengers, a *route* (sequence of stops) is a path taken to reach the destination from a source along the intermediate stops. Furthermore, other information such as the list of routes operated by different operators, end-to-end travel cost, frequency of operation, specific day and time (e.g., weekday, weekend, special days, peak hours, off-peak hours, day-time, night-time routes), etc., are also available in a few datasets.

Like other complex networks, the availability of huge data have posed big challenges to transport network analysis. Fortunately, the obtained datasets for PTNs are relatively midget, and can be processed in a reasonable time. Here, we describe three basic steps in mining the crude datasets for extracting meaningful information:

Step 1: Eliminate the anomalies that are commonly observed in the extracted datasets, e.g., data redundancies with respect to the locations of public transport stops or routes, missing information in the sequence of stops along a route, allocation of multiple id's to a specific stop or route, missing information on the geographical location of a few stops, etc.

Step 2: Process the crude data obtained in Step 1 to permit further analysis. This involves the following procedure:

(i) Since PTNs belong to the category of spatial networks, understanding the topological behavior along with spatial information will facilitate better network analysis. The spatial information of public transport stops listed in the datasets are either easting-northing or latitude-longitude. However, since many of the network visualization tools adopts latitude and longitude information for displaying the spatial locations of the stations, it is useful to convert easting and northing data to latitude and longitude using tools like ArcGIS [8]. Before the conversion, a suitable global coordinate system (e.g., WGS84) should be chosen based on the information about local coordinate systems (e.g., OSGB36 for London and HK1980 for Hong Kong) provided by the local survey departments [9].

(ii) In some datasets, the numbers assigned for the stops are typically non-sequential in nature, posing computational challenges to analysis, e.g., in generating adjacency matrix as discussed in Step 3. Thus, it is necessary to map the list of id's (both routes and stops) extracted from the database with sequentially mapped numbers. This mapping of original stop id's with sequentially mapped id's makes it less arduous to further process the data.

(iii) The concept of short distance station pairs (SSPs) has been commonly adopted to represent a group of stations as a single (merged) station [10], [11], [12]. Assigning new id's to such SSPs according to the sequential mapping carried out in Step 2 is recommended to facilitate identification of SSPs in a network. The clustering of multiple stations into one station can be based on geographical closeness, similar names for nearby stations, stations within a specific walkable catchment, etc. Although different terminologies have been used, the essential idea of SSPs has been reported in several sources [10], [11], [13]. The idea behind identifying SSPs is to establish a virtual connectivity among the nodes, especially when a large number of SSPs are observed in the network [10], [13]. However, as discussed in ref. [13], when combining multiple nodes as a single node based on their geographical closeness, the actual definition of geographical closeness is always a matter of choice. A distance threshold ($d_{\rm th}$) is needed to define the closeness of two nodes and can be chosen judiciously by observing the distribution pattern of geographical distances between successive stations (d_{ij}) in a network. However, it should be noted that choosing an extremely small value of $d_{\rm th}$ creates a lot of SSPs in a dense network, whereas a large value of $d_{\rm th}$ is meaningless, since a long walking distance to reach another station in the network is unreasonable. In either of the cases, the chosen value of $d_{\rm th}$ may bias the understanding of network behavior [13]. Hence, a careful selection of the $d_{\rm th}$ is important. SSPs are more prevalently observed in bus transport networks as compared to metro transport networks.

Step 3: Generate the topology of a PTN from the data extracted in Step 2. Initially, based on the graph type and the space of representation, a square adjacency matrix A with dimension $N \times N$ and elements a_{ij} can be derived to describe the connection between node pair n_i and n_j. The element a_{ij} = 1 if there exists a connection between nodes n_i and n_j, and 0 otherwise. A graph can either be directed (digraph), undirected, weighted or unweighted. The intent of choosing the graph type solely depends on the necessity of the type of analysis to be accomplished. For the analysis of transport structures, especially bus transport structures, a directed graph is often chosen since the inbound and outbound routes have different travel paths servicing different stations (except the round-trip journey routes). However, an undirected graph is typically chosen in the analysis of metro transport networks where the inbound and outbound travel paths remain the same for a vast majority of routes. Furthermore, depending on the aim of the network analysis, the graph can be represented in various spaces of representation as will be discussed in Section III. Thus, the type of graphs (directed, undirected, weighted, unweighted) along with the space of representation (L-, P-, B- and C-space) defines the topology of a PTN structure to be examined. Table I shows the graph type and the space of representation chosen in various PTN analysis in the literature.

Finally, for visualizing a network, there are many open source network visualization tools, and the selection would depend on the need of the analysis. For a comparison of different visualization tools, interested readers are referred to ref. [14].

III. Spaces of Network Representation

In this section, we describe different spaces of network representation together with the adjacency matrix representation for analysis of public transport networks. Our discussion will follow the basics introduced in Kurant and Thiran [15] and Ferber *et al.* [17] for representing a public transport network in different spaces of network representation, as shown in Fig. 1. The various topological representations are fundamentally related to how the network and its parameters are being perceived. For instance, different aspects of interest may include information about the stations having more routes traversing through them, the most significant station in a network in terms of connectivity, the routes servicing more stations, edges with more overlapped routes, the number of transfers needed to reach two different stations in a network, and so on. Fig. 1 shows the most commonly used representations of a PTN analysis along with their adjacency matrix entries.

(i) A graph in L-space, also called the space-of-stations, is shown in Fig. 1(b). In an L-space graph, a public transport stop is treated as a node, and a pair of nodes are connected by an edge if there is at least one route servicing the two stops consecutively. The L-space representation is the most extensively used representation in the analysis of PTNs since it signifies the actual physical infrastructure that exists in a real-world network, and renders useful information on relationship between the nodes.

(ii) A graph in B-space, also called a bipartite graph, is shown in Fig. 1(c), where both the routes and stops are represented by nodes. A route node is connected to all the stops it services, and a stop node is connected to all the routes servicing it. There is no directed

TABLE I: Graph type and space of representation used in various PTN analyses.

	Directed	Undirected	Weighted	Unweighted	References
	Bus transport network				
L-space	✓	•	•	✓	[15] [16] [17] [18] [11]
	✓	•	✓	•	[12] [19] [20] [21] [10] [22] [13]
P-space	✓	•	•	✓	[23] [16] [17] [24] [18] [25]
	✓	•	✓	•	[26] [20] [22] [10]
C-space	✓	•	•	✓	[17]
	•	✓	•	✓	[27]
	Metro transport network				
L-space	•	✓	•	✓	[28] [29] [30] [31] [32]
	•	✓	✓	•	[33] [34] [35]

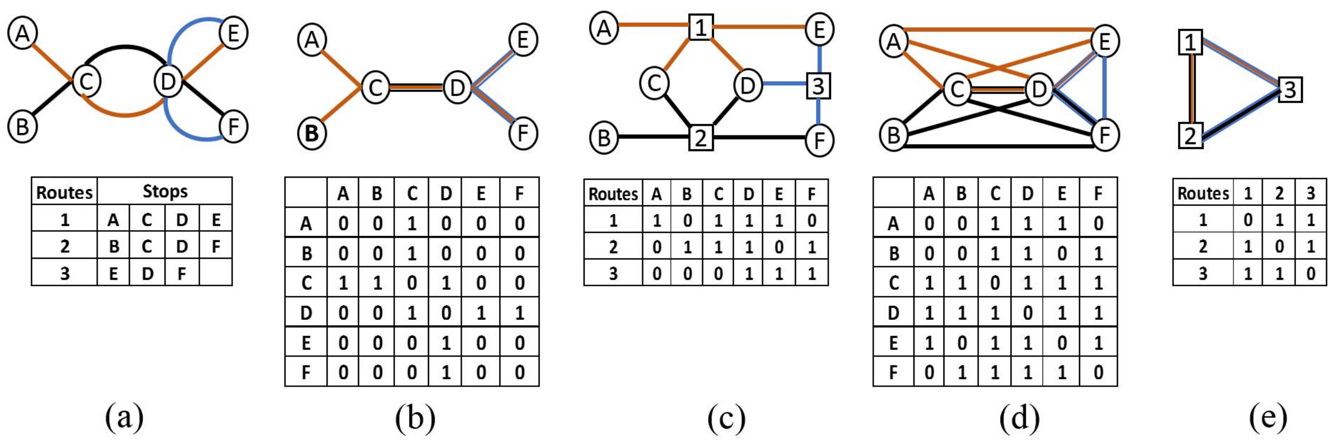


Fig. 1: (a) Simple public transport map with stations A–F being serviced by route no. 1 (shaded orange), no. 2 (black), and no. 3 (blue); (b) L-space graph; (c) B-space bipartite graph (route nodes are shown as squares); (d) P-space graph (complete sub-graph corresponding to route no. 1 is highlighted in orange); (e) C-space graph of routes. The matrix of connectivity is shown below the corresponding network representation.

edge between nodes of the same type, i.e., an edge will not exist between two route nodes or stop nodes. A graph in the B-space will be undirected. Although analysis of PTNs using bipartite graphs finds limited application, the one mode projection of a bipartite graph into the P-space (node projected) and the C-space (route projected) has gained significant attention.

(iii) A graph in P-space is also called space-of-changes, space-of-transfers, or stop-unipartite graph, and is shown in Fig. 1(d). In the P-space, the stops are represented by nodes and every possible pair of nodes that can be reached without making any transfers are linked by edges (stops serviced by a single route). A graph in the P-space can be undirected or directed depending on the type of transport networks (BTN or MTN) under study. The P-space representation renders useful information for studying the transfers between different routes since the neighbors of a node in the P-space representation are the set of nodes that can be reached with or without making a transfer. Hence, the node set associated with a specific route forms a clique or a complete subgraph.

(iv) A graph in the C-space is also called route-unipartite graph, as shown in Fig. 1(e). In the C-space, the nodes are the routes and two nodes are connected by an edge if they service a common set of stop(s) along their journeys. A graph in the C-space can be directed or undirected depending on the type of networks under study (BTN or MTN).

Table II shows the allowed graph types (directed or undirected) with respect to various spaces of network representation (L-, B-, P- and C-space) and the type of transport networks (bus or metro).

IV. Overview of Topological Analysis of Public Transport Networks

Network Science by itself has no strong association with any single field of study as its applications can be found in

TABLE II: Allowed graph type under various spaces of representation.

Space	**Directed**	**Undirected**
L	Yes	Yes
B	No	Yes
P	No	Yes
C	Yes	Yes

a great variety of real-world systems. There are a handful of parameters commonly used for analyzing complex networks. In this section, some key network parameters that aid the understanding of public transport networks are discussed. For brevity and convenience of discussion, a nomenclature list is given in Table III.

The topology of the network under analysis is represented as a graph G, which is an ordered pair comprising a set of nodes (V) and a set of edges (E), i.e., $G = (V, E)$ such that

$$V = \{n_1, n_2, n_3, ..., n_N\}; \quad N = |V| \tag{1}$$

$$E = \{e_1, e_2, e_3, ..., e_L\}; \quad e_i \to (n_i, n_j) \ \ \forall n_i, n_j \in V, \quad e_i \in E; \quad L = |E| \tag{2}$$

where N and L are the cardinality of the node set and edge set, respectively. Appendix B lists the statistical details of various PTN structures analyzed in the literature. Tables IV to VI provide an empirical comparison of a few network parameters employed in the analysis of PTNs using various spaces of representation, the details of which will be discussed in the subsequent subsections.

A. Connectivity in Public Transport Networks

In a public transport network, the connectivity pattern of a node with its neighbors is evaluated by a network parameter termed *degree,* which is the number of edges incident on a node. Degree is one of the most fundamental, yet significant parameters in network analysis. Degree is a local property of a node, and average degree of a network is a global parameter which conveys information on the average connectivity of nodes in the entire network. Depending on the graph type, the degree, k, and average degree, $\langle k \rangle$, for undirected networks are defined as

$$k_i = \sum_{j=1}^{N} a_{ij} \quad \forall i, j \in V, i \neq j, \quad \langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{3}$$

For undirected networks they are

$$k_i^{\text{in}} = \sum_{j=1}^{N} a_{ji}, \quad k_i^{\text{out}} = \sum_{i=1}^{N} a_{ij}, \quad k_i^{\text{total}} = k_i^{\text{in}} + k_i^{\text{out}} \quad \forall i, j \in V, i \neq j \tag{4}$$

$$\langle k^{\text{in}} \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^{\text{in}}, \quad \langle k^{\text{out}} \rangle = \sum_{i=1}^{N} k_i^{\text{out}}, \quad \langle k^{\text{total}} \rangle = \langle k^{\text{in}} \rangle + \langle k^{\text{out}} \rangle \tag{5}$$

All symbols in equations (3)–(5) are defined in Table 3.

The weighted node degree and the average weighted node degree are defined similar to (3)–(5), where a_{ij} is multiplied by w_{ij}, the edge weight (to be discussed in Section IV-J). Furthermore, Tables IV to VI tabulate the empirical values of average node degree under various spaces of representation. From Table IV, we observe that the average node degree in L-space analysis is nearly equal to two (in general) indicating that a stop is merely connected to its neighboring stops. On the other hand, the values shown in Table V indicate that the average node degree in the P-space analysis is roughly 10 times higher than that in the L-space which denotes the average number of nodes that can be reached from a certain node with or without making a transfer. Appendix C lists the various interpretations of the node degree under different spaces of network representation. The key point is that significant features like connectivity of a node in the L-space representation, route overlapping pattern in the C-space representation, and the number of transfers to be made in the P-space representation can be more readily identified via studying the node degree. In addition, the study of the degree distribution in a network would benefit the evaluation of an interesting network property called the *scale-free* property.

B. Are Public Transport Networks Scale-free?

Following the random network model proposed by Paul Erdős and Alfréd Rényi [38], many real-world networks were verified to be connected in a random way, in which a myriad number of nodes in the network exhibit similar degree since the nodes are connected randomly. The degree distribution of such a random network is more likely to follow a Poisson distribution [39], [38]. However, Barabási [2], [5], [40], [41] showed a unique behavior in which a few nodes in the network exhibit very high degree while a large number of nodes exhibit low degree, and the degree distribution of such network is expected to follow a power law distribution. Such networks are called scale-free networks. Observing the scale-free property in public transport networks can be inspiring since it demonstrates a strong prevalence of the hierarchical network structure, i.e., hubs at the top of the hierarchy serves maximum demand, while those below are relatively midget nodes serving mediocre demand. Intuitively, although we would expect a certain number of stops in a network are serviced by a large number of routes, it is intriguing to verify such property mathematically. Interestingly, it was observed that some of the public transport networks do exhibit the scale-free

TABLE III: Nomenclature List.

Notation	**Details**
A	adjacency matrix
a_{ij}	an element of A, defining the directed connectivity between nodes i and j
a_1	percentage of total population accessing stops on layer α
a_2	percentage of total population accessing stops on layer β
C_i	local clustering coefficient of node i
$\langle C \rangle$	average clustering coefficient
C_Δ	global clustering coefficient
$C_b(i)$	betweenness centrality of a node i
$C_b(e_{im})$	betweenness centrality of an edge e_{im} connecting nodes i and m
$C_c(i)$	closeness centrality of a node i
c_d	cost of a shortest path d
$\langle d \rangle$	average path length between two nodes
$\langle d_{tr} \rangle$	average path length between two nodes considering the number of transfers
d_{ij}	geodesic path or shortest path between nodes i and j
$d_{ij}(k)$	shortest path between nodes i and j through the node k
d_{max}	diameter of the network
d_m	# of points-of-interests of category m
$DF_{R'}$	duplication factor of a bus route R'
E or M	set of edges in a network
k_i	degree of a node i
k_{i_α}	degree of a node i on layer α
k_{max}	maximum degree of a node i
k_{min}	minimum degree of a node i
k_i^{in}	in-degree of a node i in a directed network
k_i^{out}	out-degree of a node i in a directed network
k_i^{total}	total degree of a node i in a directed network
k_i^w	weighted degree of a node i
$(k_i^{\text{in}})^w$	weighted in-degree of a node i in a directed network
$(k_i^{\text{out}})^w$	weighted out-degree of a node i in a directed network
$(k_i^{\text{total}})^w$	weighted overall degree of a node i in a directed network
$\langle k \rangle$	average degree of an undirected network
$\langle k^w \rangle$	average weighted degree of an undirected network
$\langle k^{\text{in}} \rangle$	average in-degree of a directed network
$\langle (k^{\text{in}})^w \rangle$	average weighted in-degree of a directed network
$\langle k^{\text{out}} \rangle$	average out-degree of a directed network
$\langle (k^{\text{out}})^w \rangle$	average weighted out-degree of a directed network
$\langle k^{\text{total}} \rangle$	average overall degree of a directed network
$\langle (k^{\text{total}})^w \rangle$	average weighted overall degree of a directed network
L	cardinality of edges in a network, i.e., $L = \|E\|$
L_{proj}	link projected graph of a bipartite graph
n_i^*	i^{th} node in a network
N	cardinality of nodes in a network i.e. $N = \|V\|$
N_k	number of nodes with degree k
N_{proj}	node projected graph of a bipartite graph
p_k	probability of finding a node with degree k
P_i	# of people accessing stop i
P_α	# of people accessing the stops on layer α
P_β	# of people accessing the stops on layer β
P_T	total population
R	the number of bus routes a stop joins
R'	number of routes operating between two nodes
S	the number of stops in a bus route
tr_{ij}	number of transfers between nodes i and j
V	set of nodes in a network
v_{ij}	average vehicular speed along an edge connecting nodes n_i and n_j
w_{ij}	weight of an edge connecting nodes i and j
$(w_{i_\alpha})_Z$	weight of a node i on layer α in a zone Z
w_i	overall weight of a node i
λ	Poisson coefficient
α	exponential coefficient
γ	power law coefficient
γ_{in}	power law coefficient for in-degree in a directed network
γ_{out}	power law coefficient for out-degree in a directed network
r	assortativity coefficient
$r^{(2)}$	assortativity coefficient for second neighbors of a node
σ	small-world parameter
ω	new small-world parameter
ρ_{P_α}	density of people accessing stops on layer α in a zone Z
ρ_{N_α}	density of stops layer α in a zone Z
$\frac{\rho_{P_\alpha}}{\rho_{N_\alpha}}$	node occupying probability (NOP)
ε	length difference between the two routes
λ_{th}	route length divergence threshold
γ'	transfer count difference
ξ	route transfer count divergence

* a node n_i is interchangeably represented as i in a few sections for brevity.

TABLE IV: Empirical values of various network parameters in L-space repesentation

$\langle k \rangle$	C_Δ	$\langle d \rangle$	r	References
Bus transport network				
2.48-3.03	0.055-0.161	6.83-21.52	+ve	[16]
2.88-4.59	0.09-0.15	7.13-12.56	+ve	[20]
2.1-2.4	0.0004-0.0129	28.1-50.9	•	[15]
1.18-3.59	•	6.4-52	+ve	[17]
3.13	0.142	20.03	+ ve	[18]
2.25-2.50	0.06-0.08	21.09-43.02	•	[10]
•	•	10.8-14.5	•	[12]
3.67-24.58	0.07-0.26	3.87-25.69	+ve, −ve	[11]
2.65-2.92	•	•	+ve	[19]
2.65-2.92	0.05-0.09	13.82-20.9	•	[21]
1.91-3.77	0.074-0.213	9.9-102	•	[13]
Metro transport network				
•	•	10.74-15.60	•	[32]
2-2.45	0-0.077	10-16	•	[29]
2.2	0.0018	•	•	[31]
•	0.390-0.710	•	•	[34]
2-2.4	•	6.7-19.9	•	[36]
•	•	10.13-15.02	•	[32]

TABLE V: Empirical values of various network parameters in P-space representation

$\langle k \rangle$	C_Δ	$\langle d \rangle$	r	References
Bus transport network				
33.13-90.93	0.682-0.847	1.71-2.90	+ve, −ve	[16]
41.06-94.19	0.73-0.78	2.54-2.66	+ve, −ve	[20]
24.6-102.3	0.6829-0.9095	2.3-3.7	•	[15]
4-11	•	2.2-4.7	+ve, −ve	[17]
44.60-122.89	0.716-0.819	2.84-3.45	+ve, −ve	[24]
35.84-60.24	0.57-0.68	3.15-3.46	•	[10]
44.40-92.54	0.69-0.81	2.42-3.45	•	[25]
44.46-134.65	0.73-0.78	2.53-2.89	•	[23]

TABLE VI: Empirical values of various network parameters in C-space representation

$\langle k \rangle$	C_Δ	$\langle d \rangle$	r	References
Bus transport network				
11.09-151.72	2.14-28.3	1.7-4	+ve	[17]
98.1	•	•	•	[27]

TABLE VII: Degree distribution patterns from some public transport network analyses.

L-space	P-space	C-space	References
Bus transport network			
Power law	Exponential	•	[16] [20]
Shifted power law	•	•	[18]
Power law	•	•	[13]
Power law	Shifted power law	•	[19]
Heavy tailed	Power law	•	[11] [37]
Exponential	Exponential	•	[15] [17] [10]
Exponential	•	•	[26] [12]
•	Exponential	•	[23] [24]
•	Power law	•	[25]
•	•	Exponential	[27]
Metro transport network			
Power law	•	•	[33] [34]

property. Furthermore, as explained later in this section, the degree distribution in a network is a good source of inference on the network evolution [23], [42]. Thus, the study of degree distribution has attracted enormous research interest.

The degree distribution exposes the probability of a randomly selected node in the network having a degree of k, i.e.,

$$p_k = \frac{N_k}{N} \quad \text{or} \quad N_k = Np_k \tag{6}$$

where p_k is the probability of finding a node with degree k, N_k is the number of nodes with degree k, and N is the total number of nodes in the network. Interested readers may refer to [5, Chapters 3–5] to probe further into the difference between random and scale-free networks. Table VII tabulates the degree distributions of various PTNs reported in the literature. From Table VII, we make the following observations:

(i) An exponential degree distribution in L-space indicates that connecting a newly added node with the existing nodes is more likely to be random. This is in contrary to the notion of preferential attachment where newly added nodes are connected to the already existing influential nodes in the network, making the degree distribution a power-law distribution.
(ii) An exponential degree distribution in P-space indicates that defining a new route sequence in the network is more likely to be random in order to ensure a better coverage and service rather than along the influential nodes in the network.
(iii) An exponential degree distribution in C-space indicates that defining the stops along a route node is more random than defining the stops along a route to cover the influential nodes.

Thus, the degree distribution of a network provides information on the topological evolution of the public transport network in a city [23]. Up to now, some simple network evolution models have been proposed based on fitting empirical data. However, the nature of network evolution has never been verified from the actual deployment perspective. As demonstrated by Barabási [2], the existence of hubs in a scale-free network can be a result of two phenomena, namely, growth and preferential attachment. However, the feasibility of deployment of preferential attachment in a real-world network is yet to be verified!

In our previous work [13], as part of analyzing bus transport networks, we proposed a supernode graph representation, where a supernode is a cluster of geographically closely-located nodes which satisfy the criterion $d_{th} \leq 100$ m, where d_{th} is the geographical distance between two nodes. Using the supernode representation, we analyzed the scale-free behavior for three cities, and it was very interesting to observe that the Hong Kong network plausibly exhibited the scale-free property with the supernode representation, as shown in Fig. 2. In other words, a slight modification in the topological representation permitted the exposition of an important network property which otherwise was undetected under conventional graph representation. Therefore, the effect of supernodes in analyzing the pubic transport networks should not be overlooked.

Finally, it is very interesting to observe the scale-free property (sometimes called the 80/20 rule) in public transport networks. This demonstrates the fact the a myriad number of stops carry 20% of the network load, and a countable number of stops carry 80% of the load. Public transport networks having such a property are free of any scaling applied to them. The mechanism of passenger flow in a scale-free network is an important research topic from the perspective of a transport engineer, similar to the study of information spread or disease spreading by network engineers and biologists. Another core research area of practical importance is robustness analysis which aims to study the network functionality upon removal of a certain set of target nodes. It has been shown that scale-free networks are more prone to targeted attacks, in contrary to random networks which end up merely at network fragmentation on targeted attacks.

C. Network Cohesiveness

The extent to which the immediate neighbors of a node are connected to each other is examined through a property called *clustering*, which defines the level of cohesiveness in a network. Clustering, also known as the transitivity, is a local property dealing with node level information in network theory. The cohesiveness of nodes is evaluated at local level through a parameter called *local clustering coefficient*, which is given by

$$C_i = \frac{\sum_{j,h} a_{ij} a_{ih} a_{jh}}{k_i(k_i - 1)} \tag{7}$$

for undirected networks, and

$$\begin{aligned} C_i &= \frac{\sum_j \sum_h (a_{ij} + a_{ji})(a_{jh} + a_{hj})(a_{hi} + a_{ih})}{2[k_i(k_i-1) - 2k_i^{\leftrightarrow}]}; \\ k_i{}^{\leftrightarrow} &= \sum_{i \neq j} a_{ij} a_{ji} \end{aligned} \tag{8}$$

for directed networks. At the global level, the *global clustering coefficient* is given by

$$C_\Delta = \frac{1}{N} \sum_i C_i. \tag{9}$$

Again, all symbols are defined in Table 3. For an in-depth discussion of evaluating clustering by identifying triads or cliques in different graph types, interested readers are referred to ref. [43].

The study of clustering coefficients by itself has not attracted much attention from researchers in the analysis of PTNs. However, some inspiring observations can be made from the relationship between C_i and k.

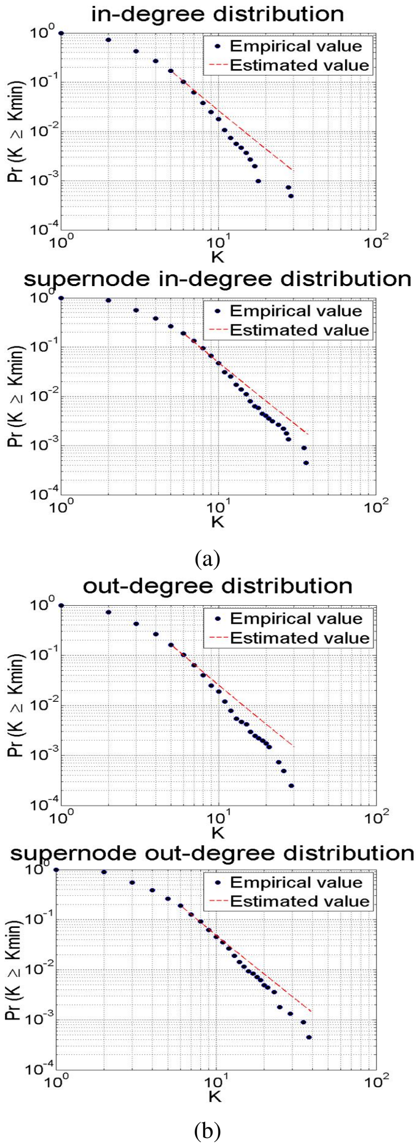


Fig. 2: Power law fit for (a) in-degree distribution; (b) out-degree distribution under regular and supernode representations.

(i) The dependency of C_i and k closely resembles a power law where the value of C_i for a given k ($C_i(k)$) is close to unity for small values of k, and $C_i(k)$ decreases rapidly with increasing k [11], [16], [20].

(ii) As observed from (7) and (8), the inverse dependency of C_i on k indicates the hierarchical structure of a network in the L-space representation, where high degree nodes (hubs) tend to form numerous connections with their neighbors, thus reducing the possibility of their neighbors having connections among themselves. This reduces the local clustering coefficient of high degree nodes. On the other hand, a low degree node has a greater tendency to be connected among its neighbors, increasing its local clustering coefficient [11].

(iii) In the P-space representation, all stations of a specific route form a perfect clique, with C_i=1 for all nodes in the route. The value of C_i becomes smaller when the nodes are shared by multiple routes. Thus, in the P-space representation, the fully connected subgraphs of all stops along a route constitute local cliques, and these local cliques are shared between routes through the common nodes. Hence, the nodes with a low degree and a high clustering coefficient belong to a fully connected local clique, whereas the nodes with a high degree and a low clustering coefficient connect multiple local cliques, reflecting that hubs act as coordinating points for several routes [16], [23], [24], [25], [44]. Thus, the distribution of $C_i(k)$ gives an indication on how the clustering is organized for nodes of various degrees.

Appendix C summarizes the common interpretations of transitivity under various spaces of network representation, and Tables IV to VI give the ranges of values of the global clustering coefficient under the various spaces of network representation. It can be seen that the clustering in P-space is significantly higher than that in L-space due to the existence of more local cliques in P-space. Although clustering has been extensively employed in L-space PTN analysis, the physical significance of evaluating both local and global clustering coefficients in L-space is vague. Moreover, the clustering coefficient is more meaningfully interpreted in the P-space representation for a PTN analysis. Also, evaluating the clustering coefficient in B-space (bipartite graph) is meaningless since the neighbors of a node are from the same group, and there exists no connection between nodes of the same group in B-space. However, evaluating clustering in C-space conveys interesting information on the extent of route overlapping in a network which is an extremely useful information for route optimization, and thus deserves more work.

D. Travel Distance in Hops

In a PTN, the number of hops to be traversed to accomplish a journey between any two chosen stops in a network is normally measured by *path length*. In graph theory, a path is a sequence of nodes connected by links. The *shortest path length* is the shortest number of links between two chosen nodes, and the *average path length* (geodesic path) is the average of the shortest path length between all node pairs in the network. The *diameter* is the longest of all shortest paths, and is an upper bound of the average path length. Although the measure of path length conveys no information on the number of transfers to be made during the journey, it is still an important measure in the public transport network analysis from a passenger point of view since the number of hops is definitely one of the prime factors considered by the passengers in selecting a route for the journey. There are a few notable algorithms for finding the average path length in a network [45]. However, it should be noted that the edge weight should be cautiously chosen (represented) in the evaluation of the average path length in a weighted graph in order to avoid a wrong interpretation of the measured path length. For example, the Dijkastra's algorithm using d_{ij} (geographical distance between two stops) and v_{ij} (average vehicular speed along a road segment) as the edge weight may generate two completely different results in evaluating the path length between two chosen nodes [46]. The average shortest path length is usually given by

$$\langle d \rangle = \frac{\sum\limits_{i \neq j} d_{ij}}{N(N-1)} \ \ \forall i = j = 1, 2, .., N \tag{10}$$

where d_{ij} is the geodesic distance between nodes n_i and n_j. Also, d_{ij}=1 if there exists a path between the two nodes, and $d_{ij} = \infty$ otherwise, implying a possible divergence problem in a non-connected graph. A smaller value of d indicates a shorter travel distance (with or without transfers) that a passenger should take to accomplish a journey. The different perspectives of average path length are given in Appendix C. A detailed comparison of average path length in different spaces has been given in Tables IV to VI. From the values of $\langle d \rangle$ given in Tables IV to VI, it is evident that the average path length in the L-space representation is significantly longer than that in the P-space representation. Thus, the average number of links traversed by a user is much larger than the number of transfers made to reach the destination. A few other notable observations concerning the average path length are

(i) An inhomogeneous distribution of stops within a city leads to Gaussian or asymmetric unimodal distribution (with longer tail ends) in the L-space and P-space representations. Thus, a fewer number of stops in the suburbs/downtown in a city leads to long travel distances. This accounts for the long tail ends in the distribution. This phenomenon is consistent with the plethora of stops observed at city centers leading to short travel distances [12], [16], [17]. A rather unique feature can be observed in the distribution pattern in ref. [17], where a secondary peak in the tail end of the distribution along with the major peak has been observed, indicating that in addition to a major central business district (CBD), a supporting minor CBD exists in the city.

(ii) As studied in ref. [15], the average path length of a network is significantly affected in L-space by the existence of shortcut paths. Despite the absence of physical connectivity between a few nodes in the PTN (e.g., between a bus stop and a metro station which are geographically close, or stops on either sides of a road segment), they can be virtually connected by a short walking distance and such nodes can be represented as short distance station pairs (SSPs) or supernodes. Thus, merely representing the physical connectivity of two different transportation networks does not justify the true measure of the average path length [10], [11], [12], [13]. However, a slight reorganization of the network topology using supernodes provides a better and more practical insight on the average path length estimation in PTN analysis [10], [13].

(iii) Fig. 3 shows the path length distribution of bus stops for the three cities analyzed in our previous work with and without considering supernodes in the network [13]. For all the three cities in Fig. 3, it has been observed that the path length values are comparatively small when the supernode representation is used which conveys more clear information on the actual path length to be traversed in practice. Thus, in a PTN analysis, the supernode representation offers a more realistic path length estimation.

(iv) The link length distribution (the distribution of geographical distance between the stops) conveys captivating information on the route length adopted by public transport networks. In ref. [10], the geographic link length distribution has been found to follow a power law, indicating that a substantial number of routes in the public transportation have a short geographical route length and only a nominal number of routes have a long route length. Furthermore, such an analysis sheds useful light on the city's demographics. (Note: Since the latitude and longitude information of the stops are given in a spherical coordinate system, the great-circle distance is preferred over the Euclidean distance in evaluating the geographic distance between two stops [47]).

(v) In PTN analysis, the average shortest path length between any two nodes in the network might not always guarantee a minimum number of transfers. Hence, combining the number of transfers with the shortest path length offers a more realistic choice for traveling between a chosen node pair. Zhang [18] has demonstrated a way of measuring the shortest path length in (10) taking into consideration the number of transfers along the shortest path, i.e.,

$$\langle d_{tr} \rangle = \frac{\sum_i \sum_j d_{ij}(1 + tr_{ij})}{N(N-1)} \ \ \forall i = j = 1, 2, ..N \tag{11}$$

where tr_{ij} is the total number of transfers needed to travel between nodes i and j.

E. Small-worldness in Public Transport Networks

First demonstrated by Watts and Strogatz [48], a class of networks, called *small-world* networks, exhibit high clustering and a low average path length. Empirically the small-world property of a network can be verified by

$$\sigma = \frac{\frac{C}{C_{\text{rand}}}}{\frac{d}{d_{\text{rand}}}} = \frac{\gamma}{\lambda} \tag{12}$$

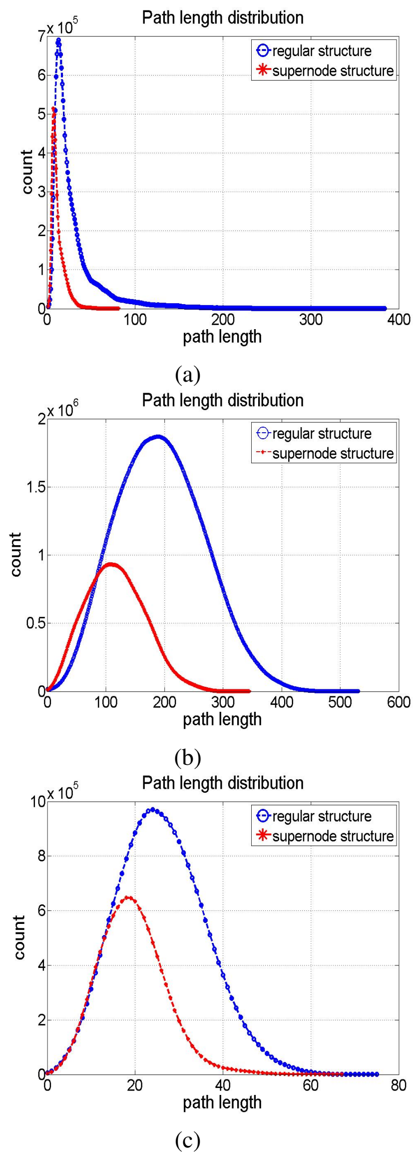


Fig. 3: Average path length distribution for (a) Hong Kong; (b) London; and (c) Bengaluru networks with and without considering supernodes.

where C_{rand} and d_{rand} are the clustering coefficient and average path length values of the equivalent random networks (degree conserved network of the same size) [49]. If $\sigma > 1$, i.e., when $C \geq C_{\text{rand}}$ and $d \approx d_{\text{rand}}$, the network can be classified as a small-world network. Telesford *et al.* [50] pointed out that the comparison of average path length of a given network to its equivalent random network is acceptable; however, the comparison of clustering of a network to that of its equivalent random network does not fully capture the small-world behavior since the clustering of a network is expected to behave close to a lattice structure. It is also observed in (12) that even a small change in C_{rand} will affect the value of the small-world parameter (σ). Hence, a new approach to capture the small-worldness of a network can be adopted, as proposed by Telesford *et al.* [50], i.e.,

$$\omega = \frac{d_{\text{rand}}}{d} - \frac{C}{C_{\text{latt}}} \tag{13}$$

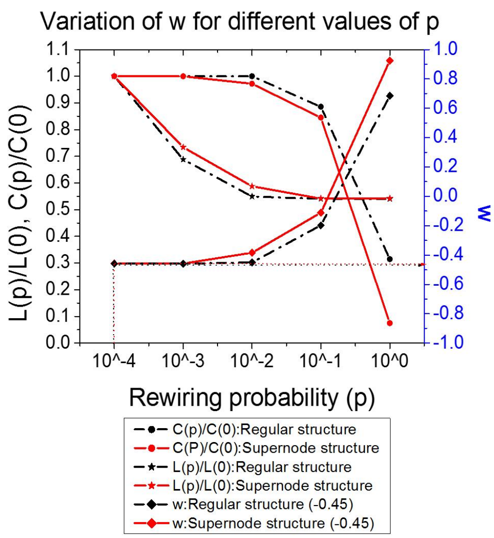
where C_{latt} and d_{rand} are the clustering coefficient and average path length values of the equivalent lattice and random network, respectively. In (13), when $C \approx C_{\text{latt}}$ and $d \approx d_{\text{rand}}$, we have $\omega \approx 0$ and such networks are considered small-world networks. By simulating the behavior of a small network, Telesford *et al.* [50] demonstrated the variation of σ and ω, where $\sigma > 1$ for all values of p (except p =1). This means that the network would show the small-world property for all the rewiring probabilities (except $p = 1$), demonstrating that $\sigma > 1$ cannot fully capture the small-worldness. However, the variation of ω shows three major zones, viz. $\omega < 0$, $\omega \approx= 0$, and $\omega > 0$, capturing the random, small-world, and lattice properties of the network [50]. Furthermore, interested readers may refer to refs. [48], [49] for details on the basic rewiring approaches.

Some reported works have attempted to use (12) to test the small-worldness of public transport networks by verifying $\sigma > 1$, but such results have been found to deliver misleading conclusions [11], [16], [24], [25], [37]. In our previous work [13], we adopted Telesford *et al.*'s method to evaluate the small-world property of bus transport networks, and the results of two networks are shown here in Fig. 4. By observing the value of ω in Fig. 4a we can see that the Hong Kong network becomes a small-world network if certain modifications are made to the existing routes. However, from the value of ω shown in Fig. 4b, we can also see that the modifications in the routes needed can be quite substantial and hence difficult to implement.

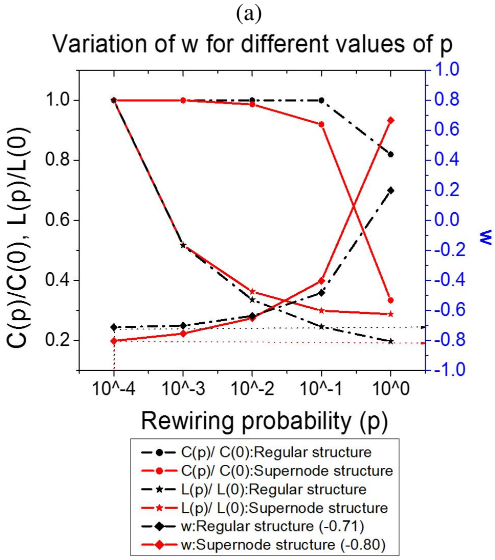
Unlike Stanley Milgram's experiment conducted in 1967 for studying the small-world behavior of a social network [51], finding a small value of average path length in large public transport networks is much more difficult. In addition, it is widely know that $\langle d \rangle$ varies with $\sqrt{N}$ [5]. Thus, a true measure of small-worldness should consider the network size as one of the parameters alongside with the clustering and average path length. Small-worldness is undoubtedly an important network behavior in public transport networks as it demonstrates the effectiveness of a transport network in terms of both connectivity (clustering) and the travel distance in hops (path length). However, existing measures of small-worldness have merely been used to demonstrate high clustering and low average path length, and a practical measure from the passenger's perspective would be more desirable for public transport networks.

F. Bridges in Public Transport Networks

Centrality is a network parameter describing primarily local information about nodes (edges), and yet having a global significance. Centrality quantifies the significance of



(a)



(b)

Fig. 4: Small world network behavior for (a) Hong Kong; and (b) London networks with and without supernodes, and the value of ω at $p = 10^{-4}$ is highlighted. (Note: L is used instead of $\langle d \rangle$ to represent the average path length).

a node (edge) based on various sources of information. Centrality measures may thus include degree centrality, Eigen-centrality, Katz-centrality, page rank centrality, closeness centrality, betweenness centrality, etc. In PTN analysis, a few centrality measures have been extensively studied, e.g., degree centrality, closeness centrality, and betweenness centrality. The degree centrality, as discussed in Section IV-A, rates a node's significance according to its degree. Similarly, betweenness centrality emphasizes the capability of a node in bridging multiple shortest paths in a network [52]. Specifically, the *node betweenness centrality* is defined as

$$C_b(i) = \sum_{i,j,k \in V} \frac{d_{jk}(i)}{d_{jk}}, \tag{14}$$

and the *edge betweenness centrality* is defined as

$$C_b(e_{im}) = \sum_{i,j,k,m \in V} \frac{d_{jk}(e_{im})}{d_{jk}} \tag{15}$$

where d_{jk} is the total number of shortest paths between nodes j and k, and $d_{jk}(i)$ or $d_{jk}(e_{im})$ is the shortest paths between nodes j and k passing through node i or edge e_{im}. Appendix C summarizes the different perspectives of betweenness centrality under various spaces of network representation.

For a given network, it is intuitive to assume that the nodes having a higher degree have a higher probability to serve as central nodes in the network, and thus, the relationship between degree and betweenness centrality has been actively studied. The major observations are as follows:

(i) The dependency of betweenness upon degree is found to follow a Poisson distribution in the L-space representation [16], and a power-law distribution in the L-space representation [22] and the C- space representations [17].
(ii) In the P-space representation, two variations of power law distribution have been observed depending on the value of k. For small values of k, the betweenness is almost zero leading to a steep slope in the power-law distribution, whereas for high values of k, a larger betweenness has been observed, leading to a more regular power-law distribution pattern [16], [17].
(iii) In the B-space representation, the distribution pattern is found to be similar to that of the P-space representation since, N_{proj} nodes have low degree and L_{proj} nodes have high degree [17]. Furthermore, Bona *et al.* demonstrated that, the nodes having a high betweenness centrality are mostly situated in CBDs [25]. However, this observation remains partially true because a node in the downtown/suburb which acts as an entry or exit point for passengers traveling between the cities might also contribute to a high betweenness centrality.

In an earlier work [53], we employed betweenness centrality as a prime parameter for studying network behavior when the interaction between multiple transport networks (bus and metro, for example) are ignored. Specifically, to demonstrate the unbalanced use or biasness of PTNs, a node weight was assigned considering the bus (layer α) and metro (layer β) transport layers as individual mono-layers where the layer interaction is ignored. Later, a method of *spatial amalgamation* was applied to integrate the two layers, and accordingly, a new node weight was assigned to the nodes in the integrated multi-layer, i.e.,

$$(P_\alpha = a_1 * P_T)_Z \quad \text{and} \quad (P_\beta = a_2 * P_T)_Z \tag{16}$$

$$(w_{i_\alpha})_Z = \left(\frac{\rho_{P_\alpha}}{\rho_{N_\alpha}}\right)_Z + k_{i_\alpha} \tag{17}$$

$$w_i = w_{i_\alpha} + C_b(i) \tag{18}$$

where all symbols are listed in Table 3. Similar to equations 16 - 18, the node weight is evaluated on layer β. Fig. 5 shows the influential nodes ($w_i \geq 0.8$) in the network according to the node weight assigned with and without considering the interaction between the layers. We can see that the assigned node weights differ significantly between the individual mono-layer analysis and the integrated multi-layer analysis. This indicates that ignoring the inter-connectedness between the transport layers leads to a unrealistic conclusions. Betweenness centrality has been employed as the prime parameter for assigning node weight for the multi-layer analysis since passengers may prefer using multiple transport networks (bus and metro) to complete their trips.

One of the main advantages of using betweenness centrality as a measure of significance of a node is that the removal of high betweenness nodes can adversely affect the average path length of the entire network as these nodes essentially control the traffic movement in the network by bridging various routes and nodes. Consideration of betweenness of nodes has recently been incorporated under robustness analysis and is attracting a significant research attention [28], [34], [54], [55], [56], [57].

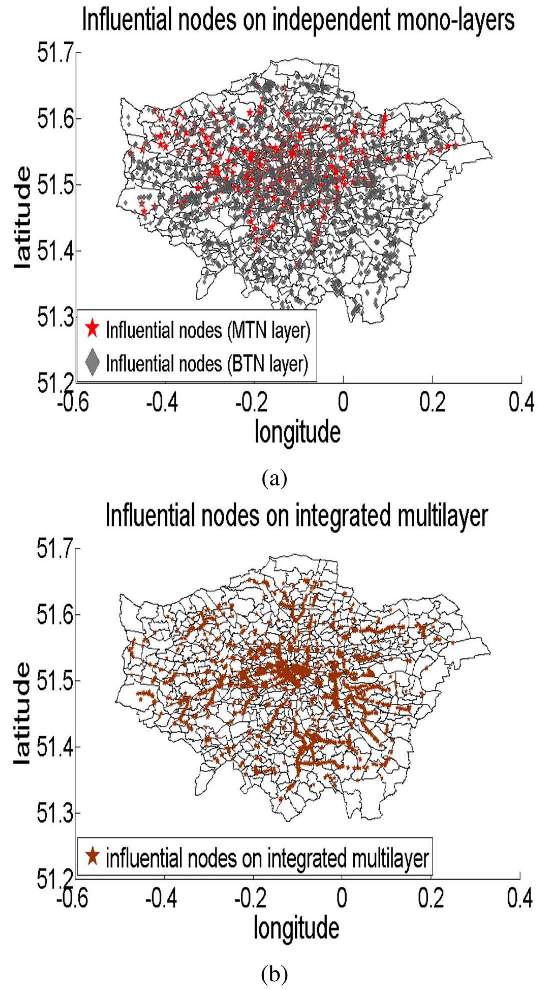


Fig. 5: Influential nodes in the London PTN with (a) mono-layer analysis; (b) multi-layer analysis.

G. *How Close are the Stops in a Public Transport Network?*

Closeness centrality is yet another parameter giving node level information, and in particular indicates how close a node i is to the rest of the network. Normally, closeness is evaluated in terms of hop count, i.e., total number of hops required to reach all other nodes in a network from a given node, i.e., we have

$$C_c(i) = \frac{1}{\sum_j d_{ij}} \tag{19}$$

The smaller the value of d_{ij}, the closer node i is to all other nodes. Prior works [12], [22] have considered the closeness centrality values for weighted and unweighted network structures, respectively, and the corresponding distributions have been found to follow an exponential distribution. Appendix C summarizes the key perspectives on closeness centrality under various spaces of representation. Due to the limited available results on closeness centrality related to PTNs and the rather restricted analysis in the L-space representation, the practical significance of evaluating closeness centrality of PTNs is still not widely recognized. In addition, in a PTN under the L-space representation, a particular stop is seldom expected to be close to all other remaining nodes in the network as it is typically connected to a portion of the network. However, closeness centrality in other spaces might offer insightful information, and should therefore deserve further investigation.

H. *Social Behavior in Public Transport Networks*

Observing the social behavior at public transport stops and routes in a PTN is interesting. Specifically, the polarization of connectivity of the stops and routes towards other stops and routes is practically useful. Such social behavior can be studied in terms of *assortativity.* While degree, as discussed in Section IV-A, captures the connectivity of a node in the network, assortativity captures the connectivity among similar kind of nodes in the network. In other words, assortativity reflects the bias of nodes to connect with nodes of similar kind. Thus, assortativity is also a local parameter providing node level information and specifically correlation between node degrees in the network. Depending on the correlation type, the network can be either assortative (connection between two high-degree or low-degree nodes) or dissassortative (connection between a high-degree node and a low-degree node). Assortativity can be assessed in terms of the average degree of a node's neighbors [58]. Moreover, Newman [59] later demonstrated that assortativity can be effectively evaluated by the Pearson correlation coefficient, i.e.,

$$r = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{k_i + j_i}{2}]^2}{M^{-1}\sum_i \frac{j_i^2 + k_i^2}{2} - [M^{-1}\sum_i \frac{k_i + j_i}{2}]^2} \tag{20}$$

where j_i and k_i are the degrees at both ends of an edge i, M is the number of edges, and $-1 \leq r \leq 1$. The network is assortative if r is +ve, and disassortative if r is −ve. Foster *et al.* [60] extended (20) for a directed network where four typical assortative mixing levels are observed, namely, r(in, in), r(in, out), r(out, in) and r(out, out) denoting the correlation between in-degree of two nodes, out-degree of two-nodes, in-degree of a node, and an out-degree of a node, respectively. The physical significance of assortativity is that a negative value of r shows the existence of core-periphery network structure and a positive value of r shows a layered network structure. In PTN analysis, it is more desirable for the network to be disassortative in order to offer better service and connectivity in a core-periphery structure. However, if a PTN follows a layered architecture, it is desirable to have assortative mixing between highly central nodes or hubs, which in turn are expected to have a disassortative mixing with other nodes in the network.

It has been observed that smaller networks are expected to be more disassortative, and larger networks exhibit both assortative and disassortative tendency [18], [21], [24]. Chatterjee *et al.* [22] developed the degree-correlation matrix to visualize the connectivity preferences of nodes in the L-space and P-space representations. Strong assortativity has been observed in L-space among low degree nodes, whereas, in P-space, strong assortativity can be seen in nodes of certain node degrees. Also, Ferber *et al.* [17] investigated the assortativity for the second neighbor ($r^{(2)}$) of a node, and found that a more positive $r^{(2)}$ indicates stronger correlation with the immediate neighbors as well as the second neighbors. Although the property of assortative mixing has so far been studied with respect to a node degree, the polarization of nodes with respect to other parameters (e.g., various centrality measures) may offer a different perspective in understanding the network behavior. Such study of social behavior of public transport stops and routes will provide important information for the design of stop locations and route distribution.

I. Communities

Community is a pair-wise parameter studied at node level and yet offers a global view in network theory. Identifying communities in a network, also called network partitioning, can be thought of as an extension to identifying assortative mixing in the network, but over a much larger set of nodes. A community is a subgraph of a network with nodes of similar behavior (in terms of connectivity), and there are dense links within a community but much fewer links between communities. Graph partitioning has been a hot research topic in the field of graph theory in the past decade since evaluating communities, especially in large and dense networks involve computationally intensive processes. An index called modularity is employed to evaluate communities in a network, as demonstrated by Newman and Girvan [61], [62], i.e.,

$$Q = \sum_i s_{ij} - \sum_{ijk} s_{ij} s_{ki} \tag{21}$$

where s_{ij} is a component of matrix s which defines the number of edges in the original network that connects nodes in community i to nodes in community j, and $0 \leq Q \leq 1$. Here, Q = 0 indicates the absence of similar degree connectivity in a network (random graph), and Q = 1 indicates a strong connection within the communities. Equation (21) has been popularly used to evaluate the modularity index for all types of networks (directed, undirected, weighted and unweighted). Moreover, in the survey conducted by Khan and Niazi [63], various modularity metrics have been considered depending on the network type. In the study by Háznagy *et al.* [12], the city's center has been found to have a few communities whereas the periphery has numerous communities. The work by Bona *et al.* [25] has identified 187 different communities with a modularity value between 0.3 to 0.7 for a PTN in a Brazilian city. For the Chinese city of Qingdao, Zhang *et al.* [19] observed a high modularity value of 0.8 with an average of 20 communities. Furthermore, a total of 46 communities with a strong modularity value of 0.91 was observed in an urban rail transit system in China [18]. Sun *et al.* [27] also found a weak modularity value of 0.34 with 7 communities in urban bus networks, where communities have been consistently identified with respect to their spatial coverage. Appendix C offers various perspectives of understanding community structures under various spaces of network representation. A physical significance of identifying communities in a network is that knowing the structural equivalence of nodes and their communities is crucial to understanding of the behavior of the intra-community and inter-community nodes.

J. Node and Edge Weights

In generating weighted networks, a weight (w) is either added to a node, an edge, or both. Weighted transport networks are still relatively less explored, despite their obvious practical significance in quantifying the relative importance of nodes and edges in relation to the level of service and performance provided by a public transport network. In this section we discuss a few weight metrics commonly employed in the topological analysis of various public transport networks.

Node weight can be assigned to reflect the relative importance of a node (station). For instance, a weight can be assigned to a station or a link according to the number of routes servicing it (degree) [12], [26], or according to the sum of weights of the adjacent edge weights (weighted degree) [27]. Edge weight may be assigned according to the morning peak hour capacity of the vehicles carrying the traffic [12], the minimum geographical distance between any two nodes [10], [21], the number of overlapped bus routes between two stations [27], [21], [11], or the number of common stops serviced along a route in C-space [20]. Furthermore, dynamic edge weights may also be assigned according to the average travel time between two nodes [19], which have been found to be very useful in analyzing the dynamic behavior of PTNs, especially in describing the varying behavior during peak- and off-peak hours.

In our recent work [13], we proposed a static demand estimation approach to assign node weight which reflects the demand centrality of a node, i.e., the capability of a node in serving the static demand by considering the number of *points of interest* (POIs), and the number of people accessing a specific station (node occupying probability). A POI can be a hospital, hotel, office, school, sports arena, cinema, shopping complex or the residential apartment. The crux of this demand estimation approach is that the real-world usage of a bus stop should be strongly dependent on the presence of POIs around the bus stop and the number of people accessing it. Using the information on POIs and node occupying probability (NOP), the node weight is evaluated as

$$w_i = c_1 \left(\sum_{m=1}^{4} d_m \right)_i + c_2 P_i + c_3 k_i \tag{22}$$

where w_i is the weight of node i, d_m is the number of POIs of category m (emergency, recreation, education, etc.) 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. Certain POIs which are equidistant to multiple stops are allocated to the nearest node with the least distance. Fig. 6 shows the heat map indicating the nodes serving high demand in Hong Kong. In a comparison between the nodes serving high demand areas and the nodes with high centrality values, we notice about 60% similarity of the nodes being compared, indicating that nodes of high topological centrality are also serving relatively higher demand areas. However, the remaining 40% nodes, though are topologically central, are serving low demand areas. This shows that merely considering topological features but ignoring their actual usage might lead to unrealistic conclusions, and such information is important information to operators to carefully design and optimize the network. Fig. 7 shows the comparison of highly central nodes versus nodes serving high demand areas. Thus, the demand estimation method would address the practical usage of topologically central nodes.

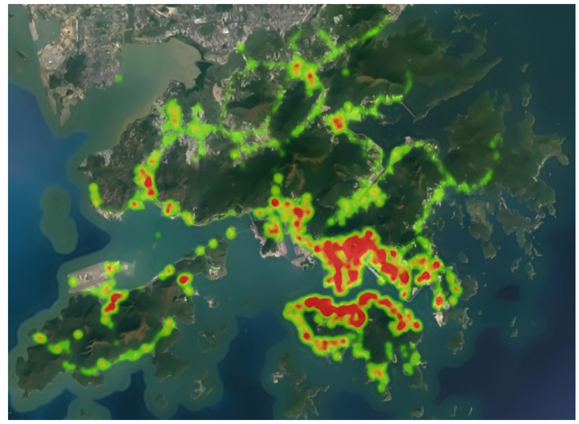
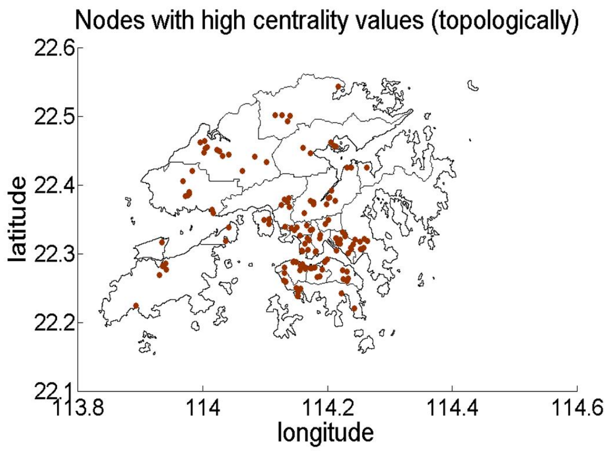


Fig. 6: Heat map showing the nodes serving higher demand areas (red) in Hong Kong.

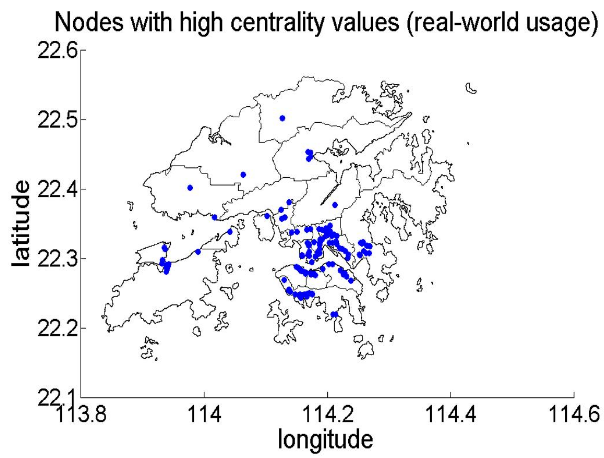
V. Notable Contributions to Public Transports Network Analysis

In this section, we discuss a few notable contributions in the field of PTN analysis in addition to the applications of network metrics in the study of PTN topologies.

(i) The usual procedure for generating the topology of a PTN is based on some available online dataset. Kurant and Thiran [15] made a novel attempt to extract real physical topology of a network by considering the timetables of the mass transportation systems. Despite the different terminologies adopted (space-of-changes for P-space representation, space-of-stations for L-space representation and the other being space-of-stops representation), the representations proposed by Kurant and Thiran [15] are generally consistent with the representation types discussed in Section III. Essentially, a multilayer framework had been adopted considering the actual mapping of logical graphs on physical graphs, where the logical layer is the real-world traffic flow layer and the physical layer is the topological representation based on space-of-changes, space-of-stations and space-of-stops. A node load was estimated based on the weighted combination of four load estimators, namely, node degree, betweenness, restricted betweenness, and simple load (origin-destination pair), assuming the combined estimation would aid in revealing some hidden network information which only degraded the performance of the best involved estimator (simple load). Moreover, Kurant and Thiran [15] also acknowledged the fact that only the OD-pair information would not suffice to carry out node load estimation without additional information like the traffic pattern.



(a)



(b)

Fig. 7: Hong Kong bus transport network with highly central nodes evaluated using (a) different centrality measures; and (b) static demand estimation method (nodes with high centrality or high node weight are the nodes with the normalized value greater than or equal to 0.8.)

(ii) A rare but insightful attempt was made by Haznagy *et al.* [12] to apply the page ranking concept in a PTN analysis. The public transport stops are ranked, in a similar manner as in web page ranking in a search engine demonstrated earlier by Larry Page [64]. The idea behind evaluating the pagerank is to identify the key nodes in the network that have significant impact in analyzing the transport efficiency.

(iii) Spatial embedding networks (SENs) have been introduced by Yang *et al.* [10] to demonstrate the effectiveness in capturing the topological properties alongside with the underlying spatial characteristics of a network. It has been demonstrated that, considering the underlying geographical feature is as important as considering the network topology in PTN analysis. A concept of extended space (ES) model was adopted to represent the L-space (ESL), P-space (ESP) and networks with SSPs (ESW) representation. A flexible transfer algorithm using the extended model was also proposed to evaluate the cost of a transfer plan (c_d) taking into account factors like transfer time, walking distance, and distance to taking buses. Such analysis has practical significance as it provides the passengers a list of top minimum cost transfer path routes.

(iv) A simple network evolution model using a quasi-continuous approximation model was proposed by Chen *et al.* [23]. In their work, the number of bus routes a stop joins, R, and the bus stop's degree, k, are the key parameters. Based primarily on the preferential attachment, a simple BTN model was organized by adding one new route at a time. It was demonstrated empirically that a strong linear correlation exists between R and k, and this formed the basis for the evolution model [23].

(v) A new P-space representation that considers the uplink and downlink routes separately for the bus routes in Harbin (a northeastern Chinese city) was proposed by Feng *et al.* [26]. Essentially, the representation introduced a duplication factor $DF_{R'}$ which is the ratio of repeated stations to unique stations for a given route R'. This parameter provides practical useful information about the bus route's spatial availability, and $DF_{R'}$ was found to exceed 36%. In the new representation, the adjacency matrix element a_{ij} is assigned a value 1 if the node is a part of both uplink and downlink routes, and 0.5 if the node is a part of either uplink or downlink route. This representation readily captures the richness of a node in terms of the degree, weighted degree, average shortest path length, and node weight (weighted degree/degree). The basis for evaluating the richness parameter is the so-called rich-club phenomenon, i.e., the correlation probability of nodes having high richness parameter (hub nodes). An exponential distribution was observed by probing the rich-club connectivity pattern, indicating that in a small portion of the network, the hub nodes are well connected. Furthermore, the evaluated node weight showed positive correlation with the corresponding degree, weighted degree, and number of routes along a node (R), indicating that the stations carrying maximum load are always well connected [26].

(vi) A simple and realistic routing algorithm called passenger intuitive logic (PIL) was used by Wu *et al.* [32] to study the passenger flow in metro networks. The passengers' intuitive strategy of choosing routes, including minimizing the number of hops traversed and the number of transfers made, formed the basis of the routing algorithm used in the study. In the study, Wu *et al.* combined the use of shortest path (SP) and minimum transfer path (MTP) to determine the routes chosen by passengers. Here, MTP corresponds to the route that has the least number of transfer times, i.e.,

$$P_{\text{MTP}} = \left(1 - \frac{\varepsilon^2}{\lambda_{th}^2}\right)^{\frac{1}{2}} \left(1 - \frac{(\gamma' - \xi)^2}{\xi^2}\right)$$
$$\text{where } \varepsilon \in [0, \lambda_{th}], \gamma' \in [0, \xi]; \text{and } P_{\text{SP}} = 1 - P_{\text{MTP}} \tag{23}$$

where P_{MTP} is the probability of taking a minimum transfer path, and P_{SP} is the probability of taking a shortest path. Simulation results for the Beijing, Tokyo, Hong Kong, and London metro systems offer insightful observations on the relationship between the topological structure of metro networks and traffic flow [32].

(vii) Topological efficiency of traffic networks has traditionally been evaluated in terms of the number of hops as given by

$$\eta_G = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{24}$$

where d_{ij} is the shortest distance path between nodes i and j. In a recent work [13], we proposed an alternative approach to measure the network efficiency in terms of time rather than distance since the time metric is more naturally used by passengers, i.e.,

$$\eta_{G,t} = \frac{1}{N(N-1)} \sum_{\substack{i=1\ldots n-1 \\ j=i+1\ldots n}} \frac{d_{ij}}{v_{ij}} \tag{25}$$

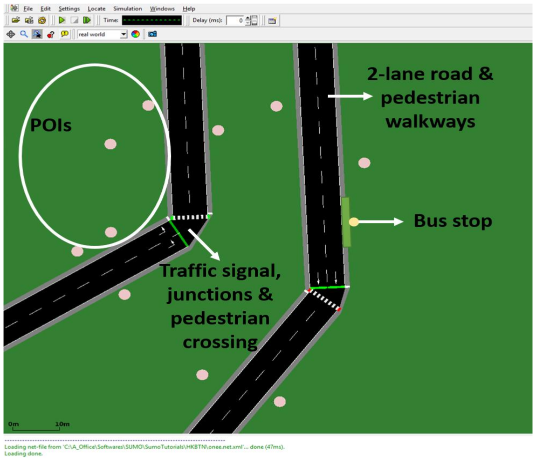
where d_{ij} is the total number of hops between nodes i and j, N is the network size, and v_{ij} is the maximum velocity attained along every hop with the shortest path between nodes i and j [13]. Due to the constraint in obtaining real-world data of v_{ij} in (25), we employed the SUMO (Simulation of Urban Mobility) simulator to validate the time efficiency for a single route using the synthetic mobility trace which yielded a better estimation of time efficiency.

SUMO is a microscopic multi-modal traffic simulator which allows the user to explicitly control the behavior of each vehicle [65]. To conduct the simulation, we first build the road network topology by importing the actual road topology including information on road junctions, bus stops, POIs, and traffic lights according to Openstreetmap [66], as shown in Fig. 8a. Then, the routes for buses are set up according to the actual time table, and other generic vehicles are set up using Activitygen, an activity-based traffic generator [67]. Finally, results are extracted from the SUMO output files which record the footprints of every vehicle during the simulation time at a sampling rate of 1 sec. By evaluating the maximum speed for every road segment and the geographic distance between the stops, the end-to-end travel delay can be calculated using (25).

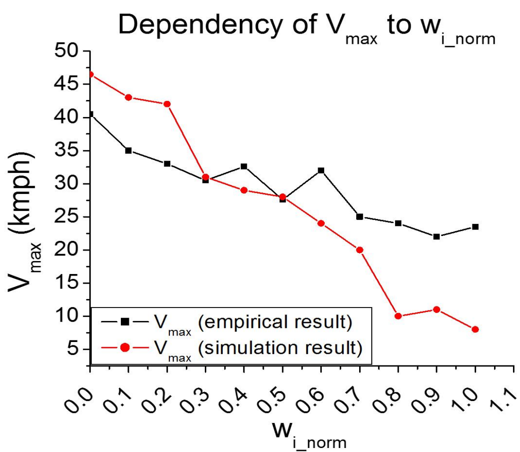
Our results showed the dependency of the vehicular speed along a road segment upon the node weight, as discussed in Section IV-J. Specifically, we observed that the higher the node weight, the lower the maximum speed attained by the vehicles on the road segment, especially during rush hours. The speed was observed to be further affected when the distance between the stops is reduced. Our simulation results have been verified using real-world data provided by the Kowloon Motor Bus Co., one of the major transport operators in Hong Kong [13]. Fig. 8b shows the dependency of the maximum speed attained along a road segment (V_{max}) for a normalized node weight (w_{i_norm}). Empirical data are in good agreement with our simulation results. We may conclude that with increased node weight (demand) and reduced geographical distance between the stops, the attainable speed by vehicles along a road segment is reduced significantly. In practice, when the bus stops are located closer to each other to offer better service, traffic speed will be compromised, and more aggressive reduction of distance between the stops may even lead to a state of traffic congestion. Our node weight model can be adopted to facilitate a better route planning and stop deployment to maintain optimal traffic performance.

VI. Conclusion and Future Work

In a data driven world, the availability of real-world datasets and high-end tools for handling huge datasets has greatly facilitated the research of complex system and data analysis. Extracting useful information from huge and distributed datasets remains a major challenge. In public transport network (PTN) analysis, the size of datasets, typically consisting of several thousand nodes, is relatively midget and the time for data mining is also comparatively manageable. Despite the successful attempts in applying concepts from network science to PTN analysis, serious study of PTN from a network science perspective is still relatively rare. In this paper, we aimed at bringing together some of the recent developments in the application of network theory to PTN analysis. In particular, useful contributions have been made by various researchers in the use of L-space representation in comparison to P-, B- and C-space representations, since the L-space graph structure mimics the actual real-world infrastructure of a PTN. A directed and weighted network structure is best suited for the study of bus transport structures, whereas an undirected and weighted network structure is more suited for metro transport studies, and the main reason for considering the graph type is the level of overlapping among inbound and outbound routes. We have found that the notion of supernodes offers practical and more insightful perspective to understanding the actual network behavior, which is difficult to be captured by conventional graph representations. Furthermore, adding static weights to



(a)



(b)

Fig. 8: (a) Snapshot of the SUMO simulator; and (b) comparison of the simulation and empirical results for the dependency of vehicular speed on node weight.

nodes and edges has been found to be effective in capturing the significance of nodes and links in PTNs. It is worth noting that merely representing the PTN structure as a graph and analyzing various network parameters may not lead to practically useful conclusions because the purpose of the public transport systems is to meet travel needs of the community being served, which requires the consideration of more practical network parameters. Also, considering the spatial embedding of PTNs alongside with the topological analysis conveys more insightful information without which quantifying the network might yield rudimentary results.

Topological analysis of PTNs have been performed using various local metrics (e.g., degree, clustering, betweenness centrality, closeness centrality), global metrics (e.g., degree distribution, scale-free property, average path length, small-world property), and pairwise properties (e.g., assortativity and communities). The study of various local, global, and pairwise properties has provided intriguing information about the topological behavior of public transport networks. Such study has provided a great source of information for researchers in the applied fields, for example, in designing of transfer algorithm, optimization of public transport routes, prediction and regulation of road congestion, network planning, transit operation, etc. However, while PTN analysis generates information like the existence of hierarchical structure, core-periphery structure, and the absence of scaling in a PTN, such information does not find immediate practical relevance to the PTN operators or government agencies. Thus, more work is still needed in developing application-oriented network analysis so that results produced from network theory can be readily translated to useful practical information and more desirably at the operational level.

Robustness analysis is another important area. Evaluating the resilience of PTNs improves understanding of various criteria of network breakdown under different attack strategies. Future research topics may also include the study of the passenger migration process, the application of integrated multiple transport modal analysis to analyze real-world complexity of passenger route selection, effects of polarization of stops and routes on the demand flow in the network, etc. Furthermore, while research efforts have been devoted to the spatial dynamics of PTNs in the past, the temporal dynamics reflecting the topological variation of a PTN at different times of the day should deserve serious attention. Another major area of research is dealing with the integration of the multiple transport networks to form a coordinated and complimentary transport system that can significantly enhance the traffic carrying capacity and efficiency of the entire system. In the past, very little contribution has been made through multi-layer analysis where individual transport networks are treated as independent topologies, and understanding the interaction among these layers should deserve more research attention in view of the practical relevance of integrated PTNs.

Finally, we would like to emphasize that applying graph theory to the analysis of public transport behavior offers an effective and convenient way to understand the network operation at both the local and global levels. The various spaces of network representation provide the fundamental network representation framework for analyzing PTNs. The incorporation of practical network parameters and the emphasis of the dynamic spatio-temporal behavior of the network can offer a broader and more practical view of the network functionality relevant to the network operators. Alongside with offering these advantages, the network-based analysis also raises a few technical challenges as a consequence of increased computational time with increasing network size and the lack of real-world datasets. In closing, we believe that PTN analysis from a graph theory perspective will continue to uncover important network properties and to serve as a solid foundation on which to develop performance optimzziation strategies, network planning, service deployment, maintenance schedules, etc. for achieving better and more sustainable transport services and eventually smarter cities.

Acknowledgment

We would like to thank Prof. John W. Polak at Imperial College London for his insightful suggestions on organizing this survey, and for shedding light on the areas of future work. We also thank Xingtang Wu at Beijing Jiaotong University for his contribution towards sharing information on metro network analysis. This work was supported in part by the National Natural Science Foundation of China (Project 61401384) and the Hong Kong Polytechnic University (Project 4-BCCH).

References

[1] M. Kuby, S. Tierney, T. Roberts, and C. Upchurch, “A comparison of geographic information systems, complex networks, and other models for analyzing transportation network topologies,” pp. 1–58, 2005.
[2] A.-L. Barabási, *Linked: The New Science of Networks*. Perseus Books Group, 2003.
[3] L. Euler, “Leonhard Euler and the Königsberg bridges,” *Scientific American*, vol. 189, no. 1, pp. 66–72, 1953.
[4] K. T. S. Oldham, *The Doctrine of Description: Gustav Kirchhoff, Classical Physics, and the “Purpose of All Science” in 19 th Century Germany*. University of California, Berkeley, 2008.
[5] A.-L. Barabási, *Network Science*. Cambridge Uuniversity Press, 2016.
[6] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
[7] M. Van Steen, *Graph Theory and Complex Networks: An Introduction*. VU Amsterdam, 2010.
[8] J. Lee and D. W. Wong, *Statistical Analysis with ArcView GIS*. John Wiley & Sons, 2001.
[9] H. J. Miller and S.-L. Shaw, *Geographic Information Systems for Transportation: Principles and Applications*. Oxford University Press, 2001.
[10] X.-H. Yang, G. Chen, S.-Y. Chen, W.-L. Wang, and L. Wang, “Study on some bus transport networks in China with considering spatial characteristics,” *Transportation Research Part A: Policy and Practice*, vol. 69, pp. 1–10, 2014.
[11] A. Chatterjee, M. Manohar, and G. Ramadurai, “Statistical analysis of bus networks in india,” *PloS One*, vol. 11, no. 12, pp. 1–17, 2016.
[12] A. Háznagy, I. Fi, A. London, and T. Németh, “Complex network analysis of public transportation networks: A comprehensive study,” in *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference on*, pp. 371–378, IEEE, 2015.
[13] T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse, “Spatial analysis of bus transport networks using network theory,” *Physica A: Statistical Mechanics and its Applications*, vol. 502, pp. 295–314, 2018.
[14] G. A. Pavlopoulos, D. Paez-Espino, N. C. Kyrpides, and I. Iliopoulos, “Empirical comparison of visualization tools for larger-scale network analysis,” *Advances in bioinformatics*, vol. 2017, 2017.
[15] M. Kurant and P. Thiran, “Extraction and analysis of traffic and topologies of transportation networks,” *Physical Review E*, vol. 74, no. 3, pp. 1–10, 2006.
[16] J. Sienkiewicz and J. A. Hołyst, “Statistical analysis of 22 public transport networks in poland,” *Physical Review E*, vol. 72, no. 4, pp. 1–11, 2005.
[17] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, “Public transport networks: empirical analysis and modeling,” *The European Physical Journal B*, vol. 68, no. 2, pp. 261–275, 2009.
[18] H. Zhang, P. Zhao, J. Gao, and X.-m. Yao, “The analysis of the properties of bus network topology in beijing basing on complex networks,” *Mathematical Problems in Engineering*, vol. 2013, 2013.
[19] X. Zhang, G. Chen, Y. Han, and M. Gao, “Modeling and analysis of bus weighted complex network in qingdao city based on dynamic travel time,” *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 17553–17572, 2016.
[20] X. Xu, J. Hu, F. Liu, and L. Liu, “Scaling and correlations in three bus-transport networks of china,” *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 441–448, 2007.
[21] H. Zhang, “Structural analysis of bus networks using indicators of graph theory and complex network theory,” *The Open Civil Engineering Journal*, vol. 11, no. 1, 2017.
[22] A. Chatterjee, “Studies on the structure and dynamics of urban bus networks in indian cities,” *arXiv preprint arXiv:1512.05909*, 2015.
[23] Y.-Z. Chen, N. Li, and D.-R. He, “A study on some urban bus transport networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 376, pp. 747–754, 2007.
[24] X. Qing, Z. Zhenghu, X. Zhijing, W. Zhang, and T. Zheng, “Space p-based empirical research on public transport complex networks in 330 cities of China,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 1, pp. 193–198, 2013.
[25] A. De Bona, K. Fonseca, M. Rosa, R. Lüders, and M. 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*, vol. 2016, 2016.
[26] S. Feng, B. Hu, C. Nie, and X. Shen, “Empirical study on a directed and weighted bus transport network in china,” *Physica A: Statistical Mechanics and its Applications*, vol. 441, pp. 85–92, 2016.
[27] L. Sun, Y. Lu, and D.-H. Lee, “Understanding the structure of urban bus networks: The c-space representation approach,” in *Proc. COTA International Conference of Transportation Professionals*, pp. 1557–1567, 2015.
[28] Y. Yang, Y. Liu, M. Zhou, F. Li, and C. Sun, “Robustness assessment of urban rail transit based on complex network theory: A case study of the beijing subway,” *Safety Science*, vol. 79, pp. 149–162, 2015.
[29] J. Zhang, M. Zhao, H. Liu, and X. Xu, “Networked characteristics of the urban rail transit networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 6, pp. 1538–1546, 2013.
[30] S. Derrible, “Network centrality of metro systems,” *PloS One*, vol. 7, no. 7, pp. 1–10, 2012.
[31] J. Zhang, X. Xu, L. Hong, S. Wang, and Q. Fei, “Networked analysis of the shanghai subway network, in china,” *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 23-24, pp. 4562–4570, 2011.
[32] X. Wu, H. Dong, C. K. Tse, I. W. Ho, and F. C. Lau, “Analysis of metro network performance from a complex network perspective,” *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 553–563, 2018.
[33] J. Feng, X. Li, B. Mao, Q. Xu, and Y. Bai, “Weighted complex network analysis of the beijing subway system: Train and passenger flows,” *Physica A: Statistical Mechanics and its Applications*, vol. 474, pp. 213–223, 2017.
[34] S. Derrible and C. Kennedy, “The complexity and robustness of metro networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 17, pp. 3678–3691, 2010.
[35] K. Lee, W.-S. Jung, J. S. Park, and M. Choi, “Statistical analysis of the metropolitan seoul subway system: Network structure and passenger flows,” *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 24, pp. 6231–6234, 2008.
[36] P. Angeloudis and D. Fisk, “Large subway systems as complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 553–558, 2006.
[37] Y. Shi, H. Lu, C. Nie, and H. Yoshitsugu, “Analysis of public transport networks in nagoya based on complex network theory,” in *Logistics: The Emerging Frontiers of Transportation and Development in China*, pp. 2759–2766, 2009.
[38] P. Erdös and A. Rényi, “On random graphs,” *Publ. Math. (Debrecen)*, vol. 6, pp. 290–294, 1959.
[39] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Magyar Tud. Akad. Mat. Kutató Int. Közl*, vol. 5, pp. 17–61, 1960.
[40] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003.
[41] A.-L. Barabási, “Scale-free networks: a decade and beyond,” *Science*, vol. 325, no. 5939, pp. 412–413, 2009.

[42] P.-P. Zhang, K. Chen, Y. He, T. Zhou, B.-B. Su, Y. Jin, H. Chang, Y.-P. Zhou, L.-C. Sun, and B.-H. Wang, "Model and empirical study on some collaboration networks," *Physica A: Statistical Mechanics and its applications*, vol. 360, no. 2, pp. 599–616, 2006.

[43] G. Fagiolo, "Clustering in complex directed networks," *Physical Review E*, vol. 76, no. 2, pp. 1–8, 2007.

[44] C. Von Ferber, Y. Holovatch, and V. Palchykov, "Scaling in public transport networks," *arXiv preprint cond-mat/0501296*, 2005.

[45] W. contributors, "Shortest path problem — wikipedia, the free encyclopedia," 2018. [Online; accessed 3-March-2016].

[46] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[47] F. Ivis, "Calculating geographic distance: concepts and methods," in *Proc. 19th Conference of Northeast SAS User Group*, 2006.

[48] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[49] M. D. Humphries and K. Gurney, "Network "small-world-ness": a quantitative method for determining canonical network equivalence," *PloS One*, vol. 3, no. 4, pp. 1–10, 2008.

[50] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti, "The ubiquity of small-world networks," *Brain Connectivity*, vol. 1, no. 5, pp. 367–375, 2011.

[51] J. Travers and S. Milgram, "The small world problem," *Phychology Today*, vol. 1, no. 1, pp. 61–67, 1967.

[52] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1-3, pp. 1–101, 2011.

[53] T. Shanmukhappa, I. W. H. Ho, C. K. Tse, X. Wu, and H. Dong, "Multi-layer public transport network analysis," in *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1–5, 2018.

[54] M. J. Alenazi and J. P. Sterbenz, "Evaluation and comparison of several graph robustness metrics to improve network resilience," in *Proc. 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, pp. 7–13, 2015.

[55] M. J. Alenazi and J. P. Sterbenz, "Comprehensive comparison and accuracy of graph metrics in predicting network resilience," in *Proc. 11th International Conference on the Design of Reliable Communication Networks*, pp. 157–164, 2015.

[56] J. Wang, "Robustness of complex networks with the local protection strategy against cascading failures," *Safety Science*, vol. 53, pp. 219–225, 2013.

[57] J. Wu, M. Barahona, Y.-J. Tan, and H.-Z. Deng, "Spectral measure of structural robustness in complex networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1244–1252, 2011.

[58] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Physical Review Letters*, vol. 87, no. 25, pp. 1–4, 2001.

[59] M. E. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, p. 208701, 2002.

[60] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski, "Edge direction and the structure of networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, pp. 10815–10820, 2010.

[61] M. E. Newman, "Detecting community structure in networks," *The European Physical Journal B*, vol. 38, no. 2, pp. 321–330, 2004.

[62] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, pp. 1–15, 2004.

[63] B. S. Khan and M. A. Niazi, "Network community detection: A review and visual survey," *arXiv preprint arXiv:1708.00977*, 2017.

[64] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," tech. rep., Stanford InfoLab, 1999.

[65] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility: an overview," in *Proc. Third International Conference on Advances in System Simulation*, ThinkMind, 2011.

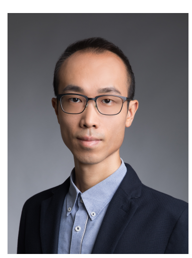
[66] "Openstreetmap." https://www.openstreetmap.org/#map=5/51.500/-0.100/. [Online. Accessed 13-Nov-2015].

[67] "Activitygen." http://sumo.dlr.de/wiki/Demand/Activity-based_Demand_Generation/. [Online. Accessed 13-Nov-2015].

[68] T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse, "Bus transport network in Hong Kong: Scale-free or not?," in *Proc. International Symposium on Nonlinear Theory and Its Applications*, pp. 610–613, 2016.

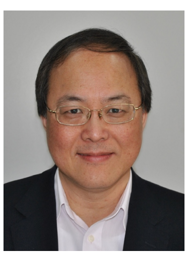
[69] V. Latora and M. Marchiori, "Is the boston subway a small-world network?," *Physica A: Statistical Mechanics and its Applications*, vol. 314, no. 1-4, pp. 109–113, 2002.

Tanuja Shanmukhappa received the B.E. degree in Electronics and Communication Engineering from the Kuvempu University in 2010, and the M.Tech. degree in Telecommunication Engineering from Visveswaraya Technological University (VTU) in 2013. She was awarded with the Ph.D. degree in Electronic and Information Engineering from the Hong Kong Polytechnic University, Hong Kong, in 2019. She was the recipient of university gold medal in her master's degree in 2013. Tanuja completed her exchange program at Imperial College London, UK, in 2018. Her current research interests include graph theory and network science, transportation network analysis, and vehicular ad-hoc networks (VANET). Tanuja also had industry experience working as an intern and trainee at Infineon Technologies and NXP semiconductors, India, respectively. She is a student member of IEEE.



Ivan Wang-Hei Ho (M'10–SM'18) received the B.Eng. and M.Phil. degrees in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2004 and 2006, respectively, and the Ph.D. degree in electrical and electronic engineering from the Imperial College London, London, UK, in 2010. He was a research intern with the IBM Thomas J. Watson Research Center, Hawthorne, NY, USA, and a Postdoctoral Research Associate with the System Engineering Initiative, Imperial College London. In 2010, he co-founded P2 Mobile Technologies Ltd, where he served as the Chief Research and Development Engineer. He is currently an Assistant Professor with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong. His research interests include wireless communications and networking, specifically in vehicular networks, intelligent transportation systems, the Internet of Things, and complex network analysis of transportation and communication networks. He primarily invented the MeshRanger series wireless mesh embedded system, which received the Silver Award in Best Ubiquitous Networking at the Hong Kong ICT Awards 2012. He is currently an Associate Editor of the IEEE Access and the IEEE Transactions on Circuits and Systems II, and served as the TPC Co-Chair of the CoWPER Workshop in conjunction with IEEE SECON 2018 and the PERSIST-IoT Workshop in conjunction with ACM MobiHoc 2019.

Chi K. Tse (M'90–SM'97–F'06) received the B.Eng. (Hons.) degree with first class honors in electrical engineering and the Ph.D. degree from the University of Melbourne, Parkville, Australia, in 1987 and 1991, respectively. He is currently the Chair Professor at the Hong Kong Polytechnic University, Hong Kong, where he was the Head of the Department of Electronic and Information Engineering from 2005 to 2012. He is the author/coauthor of ten books, 20 book chapters, and more than 500 papers in research journals and conference proceedings, and holds five U.S. patents. Dr. Tse received a number of research and industry awards, including the Best Paper Award by the IEEE TRANSACTIONS ON POWER ELECTRONICS in 2001, the Best paper Award by the International Journal of Circuit Theory and Applications in 2003, two Gold Medals at the International Inventions Exhibition in Geneva in 2009 and 2013, and was awarded a number of recognitions by the academic and research communities, including honorary professorship by several Chinese and Australian universities, Chang Jiang Scholar Chair Professorship, IEEE Distinguished Lectureship, Distinguished Research Fellowship by the University of Calgary, Gledden Fellowship, and International Distinguished Professorship-at-Large by the University of Western Australia. While with the Hong Kong Polytechnic University, he received the President's Award for Outstanding Research Performance twice, the Faculty Research Grant Achievement Award twice, the Faculty Best Researcher Award, and several teaching awards. He serves and has served as the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (2016–2017), the IEEE Circuits and Systems Magazine (2012–2015), and the IEEE Circuits and Systems Society Newsletter (since 2007), an Associate Editor for three IEEE Journal/Transactions, an Editor for the International Journal of Circuit Theory and Applications, and is on the editorial boards of a few other journals. He also serves as a Panel Member of Hong Kong Research Grants Council and NSFC and a Member of several professional and government committees



Kin K. Leung (M'86–SM'93–F'01) received his B.S. degree from the Chinese University of Hong Kong in 1980, and his M.S. and Ph.D. degrees from University of California, Los Angeles, in 1982 and 1985, respectively. He joined AT&T Bell Labs in New Jersey in 1986 and worked at its successors, AT&T Labs and Lucent Technologies Bell Labs, until 2004. Since then, he has been the Tanaka Chair Professor in the Electrical and Electronic Engineering (EEE), and Computing Departments at Imperial College in London. He is the Head of Communications and Signal Processing Group in the EEE Department. His current research focuses on protocols, optimization and modeling of wireless networks and computer systems. He also works on multi-antenna and cross-layer designs for wireless networks. He received the Distinguished Member of Technical Staff Award from AT&T Bell Labs in 1994, and was a co-recipient of the Lanchester Prize Honorable Mention Award in 1997. He was elected an IEEE Fellow in 2001, received the Royal Society Wolfson Research Merits Award in 2004-09 and became a member of Academia Europaea in 2012. Along with his co-authors, he received several best paper awards, including the IEEE PIMRC 2012, ICDCS 2013 and ICC 2019. He served as a member (2009-11) and the chairman (2012-15) of the IEEE Fellow Evaluation Committee for ComSoc. He has served as a guest editor for the IEEE JSAC, IEEE Wireless Communications and the MONET journal, and as an editor for the JSAC: Wireless Series, IEEE Trans. on Wireless Communications and IEEE Trans. on Communications. Currently, he is an editor for the ACM Computing Survey and International Journal on Sensor Networks.

Appendix A
Online Sources

City	Country	Source	Ref.
		Bus transport network	
Hangzhou	China	www.hzbuda.com.cn	[23]
Chennai	India	www.mtcbus.org/	[11]
Ahmedabad	India	www.ahmedabadbrts.org/web/commuters.html	[11]
Delhi	India	delhitravelhelp.in/StopsOfBus.aspx	[11]
Hyderabad	India	www.hyderabadbusroutes.com	[11]
Kolkata	India	www.kolkataonline.in	[11]
Mumbai	India	github.com/transitmetrics/ntd/tree/master	[11]
-	Singapore	www.streetdirectory.com.sg/	-
Hong Kong	China	data.gov.hk/en-data/category/transport?organization= hk-td/	[68]
London	UK	data.london.gov.uk/dataset/tfl-bus-stop-locations-and-routes	-
Bengaluru	India	opencity.in/topic/transportation/	-
-	Australia	opendata.transport.nsw.gov.au/search/type/dataset	-
-	-	www.apta.com	[17]
		Metro transport network	
Beijing	China	www.ebeijing.gov.cn/feature_2/BeijingSubway/	[33]
Shanghai	China	service.shmetro.com/en/	[31]
Hong Kong	China	www.mtr.com.hk/en/customer/tourist/index.php	[32]
Tokyo	Japan	www.tokyometro.jp/en/subwaymap/index.html	[32]
London	UK	tfl.gov.uk/maps/track?intcmp=40400	[32]
New York	America	web.mta.info/maps/submap.html	[32]
Boston	America	mbta.com/schedules/subway	[69]
Paris	France	parisbytrain.com/paris-metro/	[32]
Seoul	Korea	www.korea4expats.com/korean-subways.php	[35]

Appendix B
Statistical Details of PTNs

City	Place	Mode	N	Routes	Spatial analysis	Space type	Ref.
Debrecen		BET	306	53			
Gyor	Hungary	B	230	43	✓	L-space	[12]
Miskolc		BT	257	35			
Pécs		B	256	55			
Szeged		BET	242	40			
Ahmedabad			1103				
Chennai			1009				[22]
Delhi	India	B	1557	•	•	L, P-space	
Hyderabad			1088				
Kolkata			518				
Mumbai			2267				
Beijing			7864	1308			
Shanghai	China	B	5931	842	✓	L, P-space	[10]
Hangzhou			2750	509			
Shanghai			9502	1641			
Beijing	China	B	9361	1714	•	P-space	[24]
Guangzhou			3891	1256			
Shenzhen			3594	884			
Dongguan			3269	346			
Chengdu			3053	505			
Foshan			2952	378			
Hangzhou			2789	688			
Tianjin			2721	552			
Suzhou			2662	341			

Beijing			3938	516			
Shanghai	China	B	2063	501	•	L-, P-space	[20]
Nanjing			1150	174			
Warsaw			1533				
Switzerland	•	BTM	1613	•	✓	L-, P-space	[15]
Europe			4853				
Pila			152				
Belchatów			174				
Jelenia Góra			194				
Opolew			205				
Torunń			243				
Olsztyn			268				
Gorzów Wlkp			269				
Bydgoszcz			276				
Radom			282				
Zielona Góra	Poland	BT	312	•	•	L, P-space	[16]
Gdynia			406				
Kielce			414				
Czestochowa			419				
Szczecin			467				
Gdańsk			493				
Wroclaw			526				
Poznań			532				
Bialystok			559				
Kraków			940				
Lódź			1023				
Warszawa			1530				
GOP			2811				
Hangzhou			827	150			
Nanjing	China	B	1764	252	•	P-space	[23]
Beijing			4199	572			
Shanghai			4374	968			
Baoding			634	52			
Jinan	China	B	883	100	•	P-space	[19]
Shijiazhuang			1299	139			
Berlin		BMTU	2992	211			
Dallas		B	5366	117			
Dusseldorf		BMT	1494	124			
Hamburg		BFMTU	8084	708			
Hong Kong		B	2024	321			
Istanbul		BMT	4043	414			
London		BMT	10937	922			
Los Angeles	•	B	44629	1881	•	L-,P-, C-space	[17]
Moscow		BEMT	3569	679			
Paris		BM	3728	251			
Rome		BT	3961	681			
Sao Paolo		B	7215	199717			
Sydney		B	1978	596			
Taipei		B	5311	389			
Curitiba	Brazil	BM	9423	615	✓	P-space	[25]
Qingdao	China	B	1758	261	✓	L, P-space	[19]
Beijing	China	B	5421	722	•	L-, P-space	[18]
Harbin	China	B	993	132	•	P-space	[26]
Singapore	•	B	4620	428	✓	C-space	[27]
Nagoya	Japan	BM	687	280	•	L-, P-, C-space	[37]

B: Bus, E: Electric trolleybus, T: Tram, M: Metro (subway), U: Urban train, F: Ferry

APPENDIX C
NETWORK PARAMETERS IN DIFFERENT SPACES OF REPRESENTATION

Parameter	L-space	P-space	C-space	B-space
degree	number of neighboring stops that a given stop is connected to	number of stops accessible from a given stop with or without making a transfer	number of overlapped routes	number of stations serviced by a route (in L_{proj} graph) or number of routes a station is connected to (in N_{proj} graph)
local clustering (transitivity)	cohesiveness among the neighbors of a node considering the physical infrastructure	cohesiveness among the neighbors of a node considering the actual connectivity	cohesiveness among the neighbors of a node considering the common stops serviced along the routes	cohesiveness between the routes and stops in a network
average path length	total number of links (hops) to be traversed between the chosen O-D	total number of transfers to be taken to travel between the chosen O-D	-	-
betweenness centrality	node significance based on the number of shortest path routes that can traverse via the given node	node significance based on the number of transfers than can be handled by the given node	-	-
closeness centrality	reachability of a node with respect to every other node in the network	reachability of a node with respect to other routes in the network considering the number of transfers	-	-
assortativity	correlation level between similar degree stops in the network	correlation level between similar degree routes in the network	correlation between similar degree routes based on their overlapping	-
communities	identifying different zones in the network based on a behavior of the stops and their connectivity	identifying different zones in a network based on the behavior of the routes	identifying different zones in the network based on the behavior of route overlapping	-