SPATIAL ANALYSIS OF PUBLIC TRANSPORT NETWORK FROM A COMPLEX NETWORK PERSPECTIVE AND ITS IMPACT ON VEHICULAR NETWORK CONNECTIVITY

TANUJA SHANMUKHAPPA

PhD

The Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

SPATIAL ANALYSIS OF PUBLIC TRANSPORT NETWORK FROM A COMPLEX NETWORK PERSPECTIVE AND ITS IMPACT ON VEHICULAR NETWORK CONNECTIVITY

TANUJA SHANMUKHAPPA

A thesis submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

October 2018

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

_____

(Signature)

Tanuja SHANMUKHAPPA

(Name of the Student)

I would like to dedicate this work to my dearest family members and friends.

Abstract

A graph, comprising a set of nodes connected by edges, is one of the simplest yet remarkably useful mathematical structures in the analysis of real-world complex systems. By representing the components of a complex system as nodes, and their interconnectivity as edges, the inherent nature encoded in the topology of a complex system can be studied. In this thesis, we emphasize on the topological analysis of two types of public transport networks (PTNs), namely, the bus and metro transport networks, using the concepts of graph theory. The topological analysis is accomplished for three cities: Hong Kong, London, and Bengaluru. Initially, we study the bus transport network topologies of the three cities as non-interacting mono-layers. Then, we study the bus and metro network topologies as multi-layer networks by considering and ignoring the interactions between the two layers. The topological analysis unveils intriguing network behaviors under both mono- and multi-layer analyses. Lastly, we integrate the concepts from graph theory and vehicular networks to demonstrate the dependency of vehicular network connectivity on the underlying transport network topology, from a macroscopic perspective. This thesis aims to analyze the topological properties of public transport networks and their impact on the vehicular network connectivity.

First, graph theory concepts are employed to represent the PTN topology as a graph. The consideration of spatial embedding of PTNs results in a new network element called *supernode*, which yields in a novel approach in modeling the PTN topology called *supernode graph structure* representation. A *static demand estimation* approach is proposed to weigh a node's significance based on its real-world usage alongside its topological centrality. The crux of the demand estimation approach is that the accessibility factor of a node is greatly influenced by the presence of points-of-interests and the number of people accessing it. A zone classification approach is proposed which together with the static demand estimation approach aids in better identifying a node's significance.

For the part of mono-layer analysis, a few topological metrics and structural behaviors of the three bus transport networks are studied in both the conventional and supernode representations, which reveals interesting topological information from both micro- and macroscopic points of view. At every stage of the analysis, we observe that the supernode representation offers better understanding of the inherent network behaviors as compared to the conventional representation.

For the part of multi-layer analysis, initially, the bus and metro transport network topologies are considered as non-interacting mono-layers. Later, the approach of *spatial amalgamation* is used to integrate the two transport network topologies so that they can be studied as an interdependent multi-layer network. The static demand estimation approach is used to rank a node's significance with and without considering the interdependency of the multiple layers. The results demonstrate the fact that ignoring the interdependency between multiple transport modes will dramatically influence the understanding of true network behavior.

Finally, we aim to integrate the two different fields of study: network theory and VANET (Vehicular Ad-hoc Networks), to study the influence of underlying transport network topology on the vehicular network connectivity from a macroscopic standpoint. Contact duration, a prime metric in VANET is used to understand the nature of V2I (Vehicle-to-Infrastructure) connectivity, which is primarily determined by the two topological metrics: weighted in-degree and node weight. Lastly, we assess the topological robustness of the vehicular network connectivity to determine the network's ability to sustain malfunctions while continuing to offer the best possible service.

Publications

Journal papers

1. 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.

2. T. Shanmukhappa, I. W.-H. Ho, C. K. Tse, and K. K. Leung, "Recent Development in Public Transport Network Analysis From the Complex Network Perspective," *IEEE Circuits and Systems Magazine.* (under review)

Conference papers

1. 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 (NOLTA),* pp. 610-613, 2016.

2. T. Shanmukhappa, I. W.-H. Ho, C. K. Tse, X. Wu, and H. Dong, "Multilayer public transport network analysis," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 1-5, 2018.

Acknowledgments

As quoted by Albert Einstein, "It is the supreme art of the teacher to awaken joy in creative expression and knowledge". For I have experienced that joy, I would like to express my sincere and heartfelt gratitude to my supervisor Dr. Ivan Wang Hei Ho, for his continuous support, timely guidance, and invaluable advice. Personally, I think Dr. Ho is a person with great enthusiasm and keen intellect. Without his help, I could not have accomplished this PhD study.

The capacity to learn is a gift, the ability to learn is a skill, and the willingness to learn is a choice, Brian Herbert. For helping me imbibe all the three qualities, I would like to thank Prof. Michael Tse. The thought provoking discussions and insightful suggestions from Prof. Tse have encouraged me to learn a lot.

I acknowledge the Research Committee of the Hong Kong Polytechnic University for the financial support during the entire period of the project.

I would like to thank my parents and family members for their constant love and support in accomplishing this study. A sincere thanks to Yiran, Shamsa, Xingtang, Joseph, Mohit and Hari for providing the technical guidance and support when needed. I also would like to thank Nandeesh, Harish, Anjana, Manoj, Anushree, Harun, Parth and Ola for being a wonderful part of this journey. My colleagues, Situ, Elmer, Josyl, Xinyu, and Yohanna, I have enjoyed seeking from you people. My special thanks to Lily, Rose, King, Li, Josephine, Jerry, Eddie, Wai, Alex, Williams, Wendy, and Irena for bringing badminton closer to my heart, and Carol for being my running partner.

Contents

List of Figures

List of Tables

Nomenclature List

Notation	Details
A	adjacency matrix
a_{ij}	an element of A that defines directed connectivity between nodes i & j
B	cardinality of edges in a network, i.e., $B = \|E\|$
B_{proj}	link projected graph of a bipartite graph
C_i	local clustering coefficient of node i
C_{rand}	average clustering coefficient of a random network
C_{latt}	average clustering coefficient of a lattice network
$\langle C \rangle$ or C	average clustering coefficient
C_Δ	global clustering coefficient
$C_i(k)$	clustering coefficient for a given k
$C_b(i)$	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
$C(p)$	average clustering coefficient of a network for a given p
C_{existing}	average clustering coefficient for the existing network topology
c_1-c_5	constants
$\langle d \rangle$ or d	average path length between two nodes
d_{rand}	average path length of a random network
$\langle d_{tr} \rangle$	average path length between two nodes with number of transfers
d_{ij}	geodesic path or shortest path between nodes i and j
d_{th}	distance threshold
$d_{ij}(k)$	shortest path between nodes i and j through the node k
$d_{ij}(e_{im})$	shortest path between nodes i and j through an edge e_{im}
$d_{\max}$	diameter of the network
d_m	# of points-of-interests of category m
DF_R	duplication factor of a bus route R'

e_{ij}	a directed edge between nodes n_i and n_j
$\tilde{e}_{ij}$	a virtual edge between two transport layers
e_{n_i,sn_j}	a directed edge between a node n_i and a supernode sn_j
e_{sn_i,sn_j}	a directed edge between two supernodes sn_i and sn_j
E	set of edges in a network
E_s	set of superedges in a network
E_a	edge set associated with multiple layers
E_α	edge set on BTN layer in a multi-layer network
E_β	edge set on MTN layer in a multi-layer network
E_C	edge set of coupling layer in a multi-layer network
f_c	critical threshold
h_i	hub score
k_i	degree of a node i
$k_{i_\alpha}, k_{j_\beta}$	degree of a node i on layer α, and node j on layer β
$k_{\max}$	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_{w_i}	weighted degree of a node i
$k_{w_i}^{\text{in}}$ or k_i	weighted in-degree of a node i in a directed network
$k_{w_i}^{\text{out}}$	weighted out-degree of a node i in a directed network
$k_{w_i}^{\text{total}}$	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
$\langle k_w^{\text{in}} \rangle$	average weighted in-degree of a directed network
$\langle k^{\text{out}} \rangle$	average out-degree of a directed network
$\langle k_w^{\text{out}} \rangle$	average weighted out-degree of a directed network
$\langle k^{\text{total}} \rangle$	average overall degree of a directed network
$\langle k_w^{\text{total}} \rangle$	average weighted overall degree of a directed network
$L(p)$	average path length of a network for a given p
l	# of layers in a multi-layer network
M	multi-layer network
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	rewiring probability
$P(k)$	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
P_{MTP}	probability of taking a minimum transfer path
P_{SP}	probability of taking the shortest path
Q	Modularity index
r	assortativity coefficient
$r^{(2)}$	assortativity coefficient for the second neighbor
R	the number of bus routes a stop joins
R'	a route
R_I	communication range of an infrastructure
R_v	communication range of a vehicle
S	the number of stops in a bus route
sn_i	a supernode
$\tilde{sn}_i$	a giant supernode
tr_{ij}	number of transfers between nodes i and j
t_c	V2I contact duration
t_d	dwell time of a bus at the bus stop
u_i	authority score
V	set of nodes in a network
V_s	set of supernodes in a network
V_C	set of nodes in 500 m walkable catchment
$\tilde{V}$	node set in a multi-layer network
V_a	node set associated with multiple layers
v_{ij}	maximum vehicular speed along an edge
w_{ij}	weight of an edge connecting nodes i and j
w_i	overall weight of a node i
$(w_i)_{\text{zone}}$ or $(w_i)_z$	weight of a node i in a given zone

$(w_{i_\alpha})_z$	weight of a node i on layer α in a zone z
(w_{i_norm})	normalized weight of a node i
x	latitude
y	longitude
λ	Exponential coefficient
λ_i^m	Time headway of the buses
α, β	BTN and MTN layers in a multi-layer network
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
σ	small-world parameter
ω	new small-world parameter
ρ_P	static population density
ρ_m	POI density
ρ_N	node density
ρ_{P_α}	density of people accessing stops on layer α
ρ_{N_α}	density of nodes on layer α
$\frac{\rho_P}{\rho_N}$	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
η_G	topological efficiency in terms of travel distance in hops
$\eta_{G,t}$	topological efficiency in terms of travel time

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

Abbreviations List

Abbreviation	Full form
BL	Bengaluru
BT	Bluetooth
BTN	Bus transport network
CAN	Controller area network
CBD	Central business district
CCDF	Complimentary cumulative distribution function
DF	Duplication factor
DSRC	Dedicated short range communication
DT	Delay tolerant
ECDF	Empirical cumulative distribution function
ESL	Extended space for L-space
ESP	Extended space for P-space
ESW	Extended space with SSPs
FoT	Field operational test
GIS	Geographic information system
HK	Hong Kong
KMB	Kowloon Motor Bus Co.
IEEE	Institute of Electrical and Electronic Engineers
IVC	In-vehicle communication
KS	Kolmogorov-Smirnov
LD	London
LIN	Local Internet Network
LTE	Long term Evolution
MLE	Maximum-likelihood estimation
MTN	Metro transport network
MTP	Minimum transfer protocol
NoN	Network of Networks

NOP	Node occupying probability
PER	Probability error rate
PoI	Points-of-interest
PTN	Public transport network
RT	Real-time
SP	Shortest path
SUMO	Simulation of Urban Mobility
SSP	Short distance station pairs
VANET	Vehicular Ad-hoc Network
V2V	Vehicle-to-vehicle
V2I	Vehicle-to-infrastructure
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

Chapter 1

Introduction

1.1 Background

In this section, initially, a brief introduction on the topological analysis of public transport networks from a graph theoretic perspective is discussed, which is followed by a concise discussion on vehicular ah-hoc networks.

1.1.1 Public Transport Network Analysis

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 changing needs of the society, the 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 a series of challenges, including satisfying the ever-growing 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, and engineering.

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 in the field of network theory [2], application of graph theoretic concepts in the analysis of public transport

networks (PTNs) 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 modeling and analyzing of the 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, have found applications in statistical physics, biology, social sciences, finance, and engineering. One of the oldest instances of using graph theory dates back to the 17th Century when Leonhard Euler, a Swiss mathematician, used the concept of nodes and edges to solve a notable real-world problem, the seven bridges of Königsberg [3]. Another notable usage of graph theory was found by Gustav Kirchhoff, a German physicist, who employed nodes and edges to 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 relationships, citation networks, computer virus flow, network router analysis, chemical reactions, neural networks, transportation networks, etc. From the literature, it is evident that modeling various large real-world network structures as graphs, and analyzing their behaviors 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 tremendous amount of research interest in the past two decades [5, 6, 7].

1.1.2 Vehicular Ad-hoc Network

Vehicular Ad-Hoc Network (VANET) is a sub-class of Mobile-Ad Hoc Network (MANET) and an important aspect of Intelligent Transport System (ITS). VANET comprises of vehicles with on-board communication devices that allow exchange of information between the vehicles (vehicle-to-vehicle communication or V2V), between the vehicles and infrastructure (vehicle-to-infrastructure communication or V2I), and within the vehicles (in-vehicle communication or IVC).

Exchanging information between the vehicles has been in the mind of the researchers from early 19th century [8]. Though the early research in vehicular communication projects like Magic Motorways [9], Japanese CACS [10], and European Prometheus project [11] failed to commercialize the concept of vehicular communication, the investment from various mobile companies on the huge infrastructures, and the success of several prototypes and large scale field operational tests (FoTs) has attracted tremendous research interest from both the academia and industries. The commercially available versions like On star, BMW Assist, FleetBoard, and TomTom HD Traffic are a few examples which demonstrate the multi-disciplinary efforts towards increased automation and cooperation among vehicles to gather, process, and analyze the information that are useful for a wealth of on-road applications.

Furthermore, the commonly employed communication protocols in VANETs involve LIN (Local Internet Network), CAN (Controller Area Network), FlexRay, Ethernet, etc. for in-vehicle communication, and the access technologies like IEEE 802.11 WLAN (Wireless Local Area Network), WAVE (Wireless Access in Vehicular Environments), and DSRC (Dedicated Short Range Communication) are employed for V2X communication. The parameters such as dissemination range, throughput, latency, and PER (Probability Error Rate) are used for performance evaluation. Intermittent connectivity, highly dynamic and self-configured topology due to high mobility of the vehicles are some of the major challenges pertaining to VANETs. The ability of the modern vehicles to gather huge data from their on-board sensors and wireless devices, as well as the data communication among vehicles have facilitated the automotive industries to evolve over time.

The information exchanged in VANETs can be either real-time (RT) critical information or delay tolerant (DT) information which are employed to alleviate a numerous on-road problems like, efficient traffic management, congestion issues, emergency handling, alongside with providing on-road infotainment applications. Any kind of information exchange in VANETs demands some necessary factors such as highly sustained connectivity which offers longer link duration, an improved throughput, and reduced latency, which are challenging to achieve, and thus have attracted a significant research interest from multiple-domains.

1.2 Motivation

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 the connectivity, topological features and operations of many practical networked systems as well as the roles the 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 work in the field of public transport analysis from a graph theoretic perspective which is an active research area among researchers in transportation and logistics.

Although a public transport network can be unimodal or multi-modal, we focus on the analysis of two major types of public transportation, namely, the bus transport network (BTN) and metro transport network (MTN) since we believe that the two network types are most widely used by the public to meet their daily commuting needs. Unlike the conventional graph representation employed previously, we propose a novel approach termed supernode graph representation to explore the inherent and hidden topological behaviors. Such an analysis would offer insightful results that can be useful from both the operator and passenger perspectives.

Urban transportation systems are typically multi-modal in nature which comprises of bus, metro, tram, subway, light-rail, ferry, etc. In view of practical

relevance, the multi-modal transport networks should be treated as interacting networks or network of networks (NoN). However, a vast majority of analyses in the literature are dedicated towards analyzing the multiple transport modes as non-interacting mono-layers failing to quantify the structural and functional aspects of the network behavior wholly. That is, the results from analysis of non-interacting mono-layers holds good only when ignoring the presence of other layers in the network is justifiable.

Thus, to understand the dynamic relationship between the multiple layers, a PTN as a whole need to be modeled as an interconnected multilayer network that simultaneously characterizes the behavior of nodes and edges at multiple levels. Though the notion of multiplex networks (termed by various names in the literature) has already found its strong existence in other fields of study like sociology, biology, economics, and urban planning, the field of transport networks has attracted less research which motivates us to work in the direction. Specifically, we emphasize on finding the influential nodes in a network by ranking them based on their contribution to the multiple layers. Such an analysis can be a great source of information to multiple disciplines of study.

Urban vehicular networks are expected to be an integral part of future ITS which greatly assists in mitigating serious on-road issues like traffic management, congestion problems, and accidents. In VANETs, the vehicles with on-board wireless communication devices can exchange information which is primarily influenced by the spatio-temporal dynamics of vehicular mobility. Numerous works in the literature have demonstrated that the mobility pattern of a vehicle has a profound influence on the nature of vehicular connectivity, both V2V and V2I. However, in public transport networks, the mobility pattern of vehicles, e.g., buses follow a constrained and structured pattern which is significantly different from the other generic vehicles. The unique and partially deterministic nature of vehicular mobility in PTNs interests us to study the nature of vehicular connectivity in public transport networks. Specifically, we emphasize on V2I connectivity, where the vehicles are buses, and the infrastructures are the bus stops which are capable of exchanging information with the vehicles within a given communication range.

Thus, in the first of its kind, we aim at integrating the concepts of graph theory, and vehicular network connectivity analysis to understand the dependency

of vehicular network connectivity on the underlying transport network topology as shown in Fig. 1.1. Contact duration, an important metric in VANETs is employed as the performance evaluation metric. Such an analysis provides some fundamental results to the communication network deployers in offering a better and sustained IoT services, not only to on-road vehicles, but also to the passengers using PTNs. LinkNY project is an example that demonstrates the real-world application of such an analysis[12].

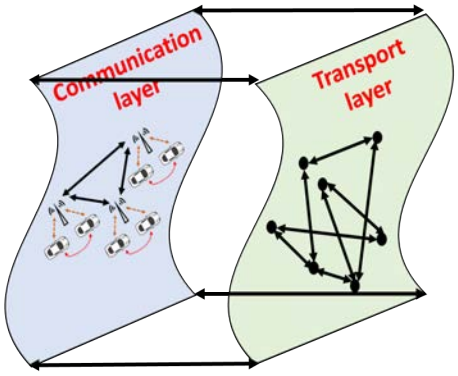


Fig. 1.1: Motivation of the work indicating the idea of integrating multiple fields of study.

1.3 Thesis Organization

This thesis is organized as follows:

Chapter 2 provides a literature review which is subdivided into two phases. The first phase emphasizes on the important network theory metrics, and their usage in analyzing the public transport networks, along with the key contributions from the literature. The second phase provides significant observations from the previous works in analyzing the connectivity dynamics in vehicular networks.

Chapter 3 discusses the basic approach of representing a PTN topology as a graph from real-world dataset using the concepts of graph theory.

Chapter 4 emphasizes on the mono-layer analysis, where, the bus transport network topological analysis is accomplished for the three cities: Hong Kong, London, and Bengaluru, from a graph theoretic perspective.

Chapter 5 discusses on the multi-layer analysis, where, the need for modeling the non-interacting transport modes as an integrated multi-layer network is demonstrated from a graph theoretic perspective.

Chapter 6 presents a high-level understanding of the impact of a PTN topology on the vehicular network connectivity. Also, the chapter discusses on the robustness of the vehicular network topology to illustrate the capability of a network to sustain malfunctions.

The thesis concludes in Chapter 7 by summarizing the major findings of the project, and some thoughts on the future work are presented.

Chapter 2

Literature Review

This chapter is divided into three sections: in the first section, we review a few works on the mono-layer analysis of PTNs from a graph theoretic perspective; in the second section, we review works on the multi-layer analysis of PTNs from a graph theoretic perspective that considers the multiple transport networks and their interdependencies; in the last section, we discuss about VANET connectivity analysis that considers contact duration as an evaluation metric.

2.1 PTN Analysis from A Graph Theoretic Perspective

Network Science by itself has no strong association with any single field of study as its applications can be found in a 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 provided at the beginning of the thesis.

The topology of the public transport network under analysis is typically 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

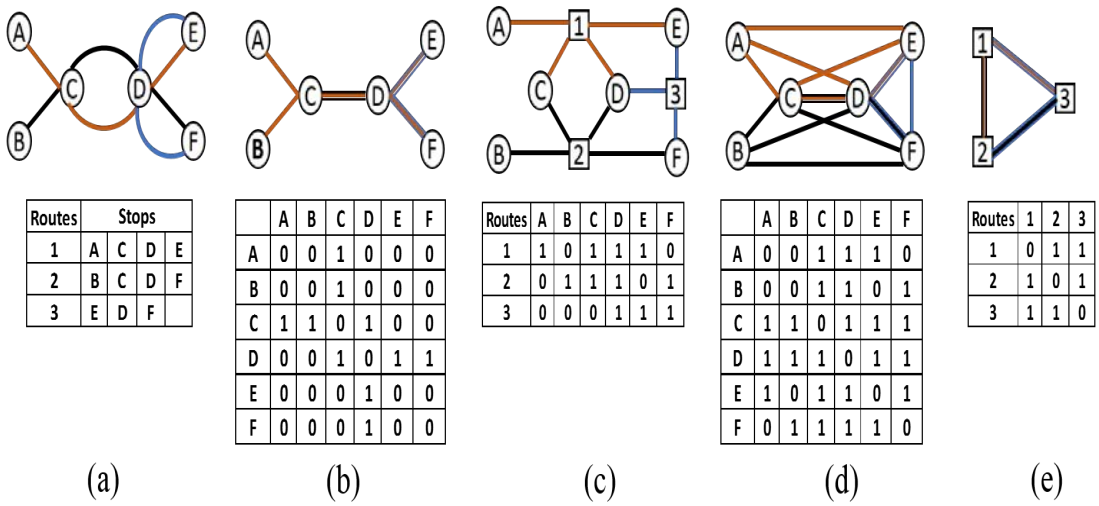
$$V = \{n_1, n_2, n_3, ..., n_N\};\ \ N = |V| \tag{2.1}$$

$$E = \{e_1, e_2, e_3, ..., e_B\};\ \ e_i \rightarrow (n_i, n_j)\ \ \forall n_i, n_j \in V,\ \ e_i \in E;\ \ B = |E| \tag{2.2}$$

where N and B are the cardinality of the node set and edge set, respectively. A graph can either be directed (digraph), undirected, weighted or unweighted. The intent of choosing the graph type solely depends on 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 2.1.1. Thus, the type of graph (directed, undirected, weighted, unweighted) along with its space-of-representation (L-, P-, B- and C-space) defines the topology of a PTN structure to be examined. Based on the graph type and the space-of-representation, a square adjacency matrix A of dimension $N \times N$ with 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.



Routes	Stops			
1	A	C	D	E
2	B	C	D	F
3	E	D	F	

	A	B	C	D	E	F
A	0	0	1	0	0	0
B	0	0	1	0	0	0
C	1	1	0	1	0	0
D	0	0	1	0	1	1
E	0	0	0	1	0	0
F	0	0	0	1	0	0

Routes	A	B	C	D	E	F
1	1	0	1	1	1	0
2	0	1	1	1	0	1
3	0	0	0	1	1	1

	A	B	C	D	E	F
A	0	0	1	1	1	0
B	0	0	1	1	0	1
C	1	1	0	1	1	1
D	1	1	1	0	1	1
E	1	0	1	1	0	1
F	0	1	1	1	1	0

Routes	1	2	3
1	0	1	1
2	1	0	1
3	1	1	0

Fig. 2.1: (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 space-of-representation.

2.1.1 Spaces of Network Representation

In this section, we describe different spaces of network representation together with adjacency matrix for the analysis of PTNs. Our discussion will follow the basics introduced in Kurant and Thiran [13] and Ferber *et al.* [14] for representing a public transport network in different spaces of network representation. 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 the more overlapped routes, the number of transfers needed to reach two different stations in a network, and so on. Fig. 2.1 shows the most commonly used spaces of representations in a PTN analysis along with their adjacency matrix entries.

1. A graph in L-space, also called the space-of-stations, is shown in Fig. 2.1(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.

2. A graph in B-space, also called a bipartite graph, is shown in Fig. 2.1(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 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.

3. A graph in P-space is also called space-of-changes, space-of-transfers, or stop-unipartite graph, and is shown in Fig. 2.1(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 are the set of nodes that can be reached with or without making a transfer since the node set associated with a specific route forms a clique or a complete subgraph.

4. A graph in the C-space is also called route-unipartite graph, as shown in Fig. 2.1(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 journey. A graph in the C-space can be directed or undirected depending on the type of transport networks under study (BTN or MTN).

Table 2.1 shows the graph type and the space-of-representation chosen in various PTN analysis in the literature. Tables 2.2 to 2.4 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.

2.1.2 Degree, Weighted Degree and Average Degree

In a public transport network, the connectivity 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 ($\langle k \rangle$) are defined as

$$k_i = \sum_{j=1}^{N} a_{ij} \quad \forall i, j \in V, i \neq j, \quad \langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{2.3}$$

Tab. 2.1: Graph type and space-of-representation used in various PTN analyses.

	Directed	**Undirected**	**Weighted**	**Unweighted**	**References**
	Bus transport network				
L-space	✓	•	•	✓	[13] [14] [15]
					[16] [17]
	✓	•	✓	•	[18] [19]
					[20] [21]
					[22] [23]
P-space	✓	•	•	✓	[14] [15] [24]
					[16] [25] [26]
	✓	•	✓	•	[20] [22]
					[23] [27]
C-space	✓	•	•	✓	[14]
	•	✓	•	✓	[28]
	Metro transport network				
L-space	•	✓	•	✓	[29] [30]
					[31] [32]
					[33]
	•	✓	✓	•	[34] [35] [36]

for undirected networks, and

$$k_i^{\text{in}} = \sum_{j=1}^{N} a_{ji}, \quad k_i^{\text{out}} = \sum_{i=1}^{N} a_{ij}, \quad k_i^{\text{total}} = k_i^{\text{in}} + k_i^{\text{out}} \quad \forall i, j \in V, i \neq j \tag{2.4}$$

$$\langle k^{\text{in}} \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^{\text{in}}, \quad \langle k^{\text{out}} \rangle = \sum_{i=1}^{N} k_i^{\text{out}}, \quad \langle k^{\text{total}} \rangle = \langle k^{\text{in}} \rangle + \langle k^{\text{out}} \rangle \tag{2.5}$$

for directed networks. All symbols in equations (2.3)–(2.5) are defined in the nomenclature list at the beginning of the thesis. The weighted node degree and the average weighted node degree are defined similar to (2.3)–(2.5), where a_{ij} is multiplied by w_{ij}, the edge weight (will be discussed in Section 2.1.11). Furthermore, Tables 2.2 to 2.4 tabulate the empirical values of average node degree under various spaces of representation. From Table 2.2, we observe that the average node degree in L-space analysis is nearly equal to two (in general) indicating

Tab. 2.2: 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	[15]
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	•	[13]
1.18-3.59	•	6.4-52	+ve	[14]
3.13	0.142	20.03	+ ve	[16]
2.25-2.50	0.06-0.08	21.09-43.02	•	[22]
•	•	10.8-14.5	•	[18]
3.67-24.58	0.07-0.26	3.87-25.69	+ve, −ve	[17]
2.65-2.92	•	•	+ve	[19]
2.65-2.92	0.05-0.09	13.82-20.9	•	[21]
Metro transport network				
•	•	10.74-15.60	•	[33]
2-2.45	0-0.077	10-16	•	[30]
2.2	0.0018	•	•	[32]
•	0.390-0.710	•	•	[35]
2-2.4	•	6.7-19.9	•	[37]
•	•	10.13-15.02	•	[33]

that a stop is merely connected to its neighboring stops. On the other hand, the values shown in Table 2.3 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 A 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. Finally, study of the degree

distribution in a network would benefit the evaluation of an interesting network property called the *scale-free* property.

Tab. 2.3: 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	[15]
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	•	[13]
4-11	•	2.2-4.7	+ve, −ve	[14]
44.60-122.89	0.716-0.819	2.84-3.45	+ve, −ve	[25]
35.84-60.24	0.57-0.68	3.15-3.46	•	[22]
44.40-92.54	0.69-0.81	2.42-3.45	•	[26]
44.46-134.65	0.73-0.78	2.53-2.89	•	[24]

Tab. 2.4: 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	[14]
98.1	•	•	•	[28]

2.1.3 Scale-free Property

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 [38, 39]. 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 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 which are serviced by a large number of routes, it is intriguing to verify such a property mathematically. Interestingly, it was observed that some of the public transport networks do exhibit scale-free property. Furthermore, as explained later in this section, the degree distribution in a network is a good source of inference on the network evolution [24, 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.,

$$P(k) = \frac{N_k}{N} \quad \text{or} \quad N_k = NP(k) \tag{2.6}$$

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. If the degree distribution of a network follows power-law, such networks are termed scale-free. The power-law is given by,

$$P\left(k\right) \propto k^{-\gamma} \Rightarrow P\left(k\right) = Ck^{-\gamma} \tag{2.7}$$

where $2 < \gamma < 3$ is the scaling parameter, k is the node degree, and C is a constant. Taking log on both sides of (2.7), we have

$$\ln P\left(k\right) \propto -\gamma \ln k \Rightarrow \ln P\left(k\right) = -\gamma \ln k + \ln C \tag{2.8}$$

Thus, the power law is a straight line in log-log scale with negative slope γ, and valid for $k \geq k_{\min}$. Interested readers may refer to [5, Chapters 3–5] to probe further into the difference between random and scale-free networks. Table 2.5 tabulates the degree distributions of various PTNs reported in the literature. From Table 2.5, we make the following observations:

1. 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.

2. 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.

3. 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 [24]. 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. But the feasibility of deployment of preferential attachment in a real-world network is yet to be verified!.

It is very interesting to observe the scale-free property (sometimes called the 80/20 rule) in public transport networks. The 80/20 rule demonstrates the fact that 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. Lastly, scale-free property is surely one of the interesting topological behaviors which has been empirically verified for three topologies in our work as will be discussed in Chapter 4.

2.1.4 Clustering or Transitivity

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

Tab. 2.5: Degree distribution patterns from some public transport network analyses.

L-space	P-space	C-space	References
Bus transport network			
Power law	Exponential	•	[15] [20]
Shifted power law	•	•	[16]
Power law	Shifted power law	•	[19]
Heavy tailed	Power law	•	[17] [43]
Exponential	Exponential	•	[13] [14] [22]
Exponential	•	•	[18] [27]
•	Exponential	•	[24] [25]
•	Power law	•	[26]
•	•	Exponential	[28]
Metro transport network			
Power law	•	•	[34] [35]

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

$$C_i = \frac{\sum_{j,h} a_{ij} a_{ih} a_{jh}}{k_i(k_i - 1)} \tag{2.9}$$

for undirected networks, and

$$C_i = \frac{\sum_j \sum_h (a_{ij} + a_{ji})(a_{jh} + a_{hj})(a_{hi} + a_{ih})}{2[k_i(k_i - 1) - 2k_i^{\leftrightarrow}]}; \quad k_i^{\leftrightarrow} = \sum_{i \neq j} a_{ij} a_{ji} \tag{2.10}$$

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

$$C_\Delta = \frac{1}{N} \sum_i C_i. \tag{2.11}$$

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. [44].

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.

1. 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 [15, 17, 20].

2. As observed from (2.10) and (2.11), 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 [17].

3. 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 [15, 24, 25, 26, 45]. Thus, the distribution of $C_i(k)$ gives an indication on how the clustering is organized for nodes of various degrees.

Appendix A summarizes the common interpretations of transitivity under various spaces of network representation, and Tables 2.2 to 2.4 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.

In the literature, 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. The usefulness of evaluating the local and global clustering of a network will be demonstrated empirically in Chapter 4.

2.1.5 Average Path Length

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 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 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 a prime factor considered by the passengers in selecting a route to accomplish their journey.

There are a few notable algorithms for finding the average path length in a network [46]. 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 [47]. The average shortest path length is usually given by

$$\langle d \rangle = \frac{\sum\limits_{i \neq j} d_{ij}}{N(N-1)} \quad \forall i = j = 1, 2, .., N \tag{2.12}$$

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 A. A detailed comparison of average path length in different spaces has been given in Tables 2.2 to 2.4. From the values of $\langle d \rangle$ given in Tables 2.2 to 2.4, 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

1. 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 [14, 15, 18]. A rather unique feature can be observed in the distribution pattern in ref. [14], 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.

2. As studied in ref. [13], 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 stop 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 average path length [17, 18, 22]. However, a slight reorganization of the network topology using supernodes provides a better and more practical

insight on the average path length estimation in a PTN analysis [22].

3. 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. [22], the geographic link length distribution has been found to follow a power law, indicating that a substantial number of routes in the public transportation has 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 [48]).

4. 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 [16] has demonstrated a way of measuring shortest path length in (2.12) taking into consideration the number of transfers along the shortest path, i.e.,

$$\langle d_{tr} \rangle = \frac{\sum_i \sum_j d_{ij}(1 + tr_{ij})}{N(N-1)} \quad \forall i = j = 1, 2, ..N \tag{2.13}$$

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

As mentioned before, the average path length is undoubtedly one of the important measures in a PTN analysis since the number of hops is definitely a prime factor considered by the passengers in deciding a route to accomplish their journey, and thus, in Chapter 4, the parametric estimation of average path length is discussed empirically for the three chosen cities.

2.1.6 Small-worldness

First demonstrated by Watts and Strogatz [49], a class of networks, called *small-world* networks, exhibit high clustering and a low average path length. Empiri-

cally the small-world property of a network can be verified by

$$\sigma = \frac{\frac{C}{C_{\text{rand}}}}{\frac{d}{d_{\text{rand}}}} \tag{2.14}$$

where C_{rand} and d_{rand} are the clustering coefficient and average path length values of the equivalent random network [50, 51]. 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.* [52] 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 (2.14) 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.* [52], i.e.,

$$\omega = \frac{d_{\text{rand}}}{d} - \frac{C}{C_{\text{latt}}} \tag{2.15}$$

where C_{latt} and d_{rand} are the clustering coefficient and average path length values of the equivalent lattice and random network, respectively. In (2.15), 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.* [52] 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 $\sigma >$ 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 [52]. Furthermore, interested readers may refer to refs. [49, 50] for details on the basic rewiring approaches. Some reported works have attempted to use (2.14) to test the small-worldness of public transport networks by verifying $\sigma > 1$, but such results have been found to deliver misleading conclusions [15, 17, 25, 26, 43].

Unlike Stanley Milgram's experiment conducted in 1967 for studying the small-world behavior of a social network [53], finding a small value of average path length in large public transport networks is much more difficult. In ad-

dition, it is widely know that $\langle d \rangle$ varies as $\sqrt{N}$ [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). Lastly, in our work, the small-worldness property will be empirically verified for the three network topologies in Chapter 4.

2.1.7 Betweenness Centrality

Centrality is a network parameter describing primarily local information about nodes (edges), and yet having a global significance. Centrality quantifies the significance of 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 2.1.2, 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 [54]. Specifically, the *node betweenness centrality* is defined as

$$C_b(i) = \sum_{i,j,k \in V} \frac{d_{jk}(i)}{d_{jk}}, \tag{2.16}$$

and the *edge betweenness centrality* is defined as

$$C_b(e_{im}) = \sum_{i,j,k,m \in V} \frac{d_{jk}(e_{im})}{d_{jk}} \tag{2.17}$$

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 A 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. It has been observed that

1. The dependency of betweenness upon degree is found to follow a Poisson distribution in the L-space representation [15], and a power-law distribution in the L-space representation [23] and the C- space representations [14].

2. 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 [14, 15].

3. 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 B_{proj} nodes have high degree [14]. Furthermore, Bona *et al.* [26] demonstrated that, the nodes having a high betweenness centrality are mostly situated in CBDs [26]. 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.

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 [29, 35, 55, 56, 57, 58]. Lastly, in our work, we will evaluate the betweenness centrality of the nodes for the three network topologies in Chapter 4, and also, in Chapter 5, we will demonstrate the need for considering the layer interdependency among multiple transport modes by employing betweenness centrality as the main parameter.

2.1.8 Closeness Centrality

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.,

$$C_c(i) = \frac{1}{\sum_j d_{ij}} \tag{2.18}$$

The smaller the value of d_{ij}, the closer node i is to all other nodes. Prior works [18, 23] 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 A 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.

2.1.9 Assortativity

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 2.1.2, 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 (connec-

tion 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 [59]. Moreover, Newman [60] later demonstrated that assortativity can be effectively evaluated by the Pearson correlation coefficient, i.e.,

$$r = \frac{B^{-1}\sum_i j_i k_i - [B^{-1}\sum_i \frac{k_i+j_i}{2}]^2}{B^{-1}\sum_i \frac{j_i^2+k_i^2}{2} - [B^{-1}\sum_i \frac{k_i+j_i}{2}]^2} \tag{2.19}$$

where j_i and k_i are the degrees at both ends of an edge i, B 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.* [61] extended (2.19) 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 [16, 21, 25]. Chatterjee *et al.* [23] 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.* [14] 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.

2.1.10 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 [62, 63], i.e.,

$$Q = \sum_{i} s_{ij} - \sum_{ijk} s_{ij} s_{ki} \tag{2.20}$$

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 (2.20) 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 [64], various modularity metrics have been considered depending on the network type.

In the study by Háznagy *et al.* [18], the city's center has been found to have a few communities whereas the periphery has numerous communities. The work by Bona *et al.* [26] 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 [16]. Sun *et al.*

[28] 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 A 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 the behavior of intra-community and inter-community nodes.

2.1.11 Node and Edge Weights

In generating weighted networks, a weight (w) is either added to a node or 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) [18, 27], or according to the sum of weights of the adjacent edge weights (weighted degree) [28]. Edge weight may be assigned according to the morning peak hour capacity of the vehicles carrying the traffic [18], the minimum geographical distance between any two nodes [21, 22], the number of overlapped bus routes between two stations [17, 21, 28], 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 work, we propose a static demand estimation approach to weigh a node's significance, and the number of routes between the two stations is employed as an edge weight, the details of which will be discussed in Chapter 3.

2.1.12 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.

1. The usual procedure for generating the topology of a PTN is based on some available online dataset. Kurant and Thiran [13] made a novel attempt to extract real physical topology of a network by considering the time-tables 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 [13] are generally consistent with the representation types discussed in Section 2.1.1. Essentially, a multilayer framework has 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 is 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 [13] 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.

2. A rare but insightful attempt was made by Haznagy *et al.* [18] 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 [65]. The idea behind evaluating the pagerank is to identify the key nodes in the network that have significant impact in analyzing the transport efficiency.

3. Spatial embedding networks (SENs) have been introduced by Yang *et al.*

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

4. A simple network evolution model using a quasi-continuous approximation model has been proposed by Chen *et al.* [24]. 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 can be organized by adding one new route at a time. It has been demonstrated empirically that a strong linear correlation exists between R and k, and this forms the basis for the evolution model [24].

5. A new P-space representation that considers the uplink and downlink routes separately for the bus routes in Harbin (a northeastern Chinese city) has been proposed by Feng *et al.* [27]. Essentially, the representation introduces 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 $DF_{R'}$ has been 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 has been 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 has 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 [27].

6. A simple and realistic routing algorithm called passenger intuitive logic (PIL) has been used by Wu *et al.* [33] 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, forms the basis of the routing algorithm that is used in the study of passenger flow in metro systems. In their 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.,

$$P_{\mathrm{MTP}} = \left(1 - \frac{\varepsilon^2}{\lambda_{th}^2}\right)^{\frac{1}{2}} \left(1 - \frac{(\gamma' - \xi)^2}{\xi^2}\right) \; \varepsilon \in [0, \lambda_{th}], \gamma' \in [0, \xi]; P_{\mathrm{SP}} = 1 - P_{\mathrm{MTP}} \tag{2.21}$$

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 [33].

Lastly, as observed from the literature, the roles the various parameters in graph theory play in determining the performance and behavior of transport systems are of great importance from both the passenger and operator perspectives. Thus, in Chapter 4 we study the PTN topological behavior from a graph theoretic perspective. Unlike the conventional way of representing a graph, in our work, we propose a novel approach which aids in more realistic topological analysis as will be discussed in Chapter 4.

2.2 Multi-layer Networks

Urban transportation systems are typically multi-modal in nature which comprises of bus-, metro-, tram-, subway-, light-rail-, ferry- networks, and so on. Considering the real-world usage of the PTNs, the multiple transport modes should be treated as interacting and interdependent networks, or network of networks (NoN), to simultaneously capture the benefaction of the nodes and edges on multiple layers. However, a vast majority of analyses in the literature are dedicated towards analyzing the isolated, independent, and non-interacting mono-layer transport networks, or the aggregated version of them, which fails to quantify the structural and functional aspects of the network behavior wholly. Thus in this section, we emphasize on the need for modeling the public transport modes as interdependent multi-layer networks by presenting the key insights from the literature.

Kivelä *et al.* [66] presented an exhaustive review of the multilayer networks by introducing a general framework for the analysis of a wide range of networks like, monoplex (mono-layer) networks, multiplex (multi-layer) networks, interdependent networks, interconnecting networks, networks of networks, etc. Tomasini [67] followed up on Kivelä's work and provided a more general framework for multi-layer network analysis. Zanin [68] discussed the multi-layer nature of the functional networks and demonstrated that neglecting the multi-layer structure has dramatic consequences in understanding the actual network behavior. Gao *et al.* [69] presented a detailed review of the analytical framework for connectivity properties in NoN which were formed by the interdependent networks.

Identifying the set of influential nodes in a network and ranking them based on their significance aids in identifying central nodes in a network which acts as a great source of information to multiple disciplines of study, which is typically accomplished by evaluating various centrality measures in mono-layer analysis. However, considering the interdependency of multiple transport modes, the significance of a node to the overall network behavior might vary which has very limited contributions in the literature. Domenico *et al.* [70] in their work capitalized on the tensorial formalism to demonstrate the natural extension of various centrality measures employed in mono-layer analysis to the realm of interconnected multi-

layer networks. The authors highlighted on the significant difference observed in ranking of nodes considering the weighted mono-layer and multi-layer networks.

Menichetti *et al.* [71] demonstrated that partial analysis of layers failed to capture the significant correlation between weights with mono- and multi- layers. The authors also introduced a framework based on entropy to quantify the information stored in multi-layer network, and observed undetected information in the mono-layers when analyzed in isolation. Chodrow *et al.* [72] considered the multi-layer aspect of a network from both structural and socio-technical aspect to analyze the nature of travel demand on the transport layer, and demonstrated that demand structure plays vital role in both qualitative and quantitative analyses of the network. Furthermore, a majority of works in the literature are dedicated towards understanding the robustness of the integrated multi-layer network by following various attack models where the nodes were removed from the network from multiple layers, and their implication on the overall network behavior was assessed [73, 74, 75].

Lastly, Existing works on PTN analysis using graph theory have emphasized on either the individual transport networks by considering them as isolated mono-layer networks, or the non-interacting aggregated multi-layer networks. Though the different transport layers share common features when analyzed as individual mono-layer structures, by understanding the interconnectedness among different mono-layers, a more meaningful insight is gained into the overall network structure and its dynamics. Also, since passengers use multiple transport modes (on different transport layers) to reach their destinations, it is of practical importance, though rarely considered, to study the interaction and connectivity between network layers of different transport modes, the details of which are discussed in Chapter 5.

2.3 VANET Connectivity Analysis

This section is divided into two sections: in the first section we discuss a few notable works on VANET connectivity analysis considering contact duration as the evaluation metric, and in the second section we briefly review a few works in the literature which emphasize on robustness analysis of a PTN topology.

Urban vehicular networks are expected to be an integral part of future intelli-

gent transport system to mitigate serious on-road issues like traffic management, congestion problems, and accidents, along with providing on-road infotainment services. The research in VANET is often dedicated to either microscopic or macroscopic analysis to understand the nature of V2V and V2I communication. In microscopic analysis, characteristics such as driver behaviors, individual vehicle mobility are taken into consideration, whereas, in macroscopic analysis road network topology, diurnal cycles, socio-technical activities are considered. Key parameters such as dissemination range, throughput, latency, link duration, inter-contact time, and re-healing period are used for performance evaluation, while vehicular density and transmission range are used in understanding the connectivity dynamics in VANETs.

In the course of analyzing the factors that influence the connectivity in vehicular network (both V2V and V2I), various mobility models [76, 77, 78], channel estimation models [79, 80], and routing strategies [81, 82] has been proposed and verified in the literature. Additionally, the spatio-temporal dynamics of vehicular movement that significantly influence the instantaneous connectivity in VANETs has been verified from a graph theoretic perspective where the topological metrics like degree, clustering, and betweenness centrality were evaluated [83, 84, 85, 86, 87, 88].

Though a numerous analytical frameworks and empirical results has been discussed in the literature to understand the connectivity dynamics in VANETs, these works mainly considered generic vehicles like cars and taxis. Very few contributions have considered different vehicle types such as public transport networks. A rather unique, and partially deterministic mobility pattern of PTNs such as fixed route and known schedules have attracted interest of a few researchers to separately model the connectivity in PTNs, specifically BTNs [89, 90].

Ho *et al.* [89] studied the interbus connectivity pattern by considering the buses in urban area, and examined their implications on transport-related services. An extensive set of simulation results and real-world data on bus routes were employed in their work. Specifically, the effect of bus stop location and the prevailing traffic patterns on vehicular connectivity were demonstrated. The authors discussed the significance of modeling the mobility of buses explicitly and the usage of such connectivity analysis in developing routing algorithms in

interbus communication networks.

Doering *et al.* [90] analyzed the realistic movement patterns of bus traces from Chicago and Seattle, and analyzed the parameters such as vehicle density, speed, update intervals, and characteristics that are specific to bus transport networks. The authors discussed that such an analysis could be useful in communication systems which forms an useful part of future smart cities.

Contact duration is a key metric in VANET connectivity analysis that strongly determines the throughput of a communication. Ho *et al.* [89] and Li *et al.* [91] are the only authors in the literature to study the nature of vehicular connectivity considering the contact duration as the performance evaluation metric. Though the vehicles considered in the two works are different, a common behavior of contact duration was observed in both the works, i.e., the dichotomy behavior. As shown in both the works, a characteristic time was observed below which the contact duration followed an exponential distribution, and above which it followed a power-law distribution.

Though the authors in [91] merely discussed the distribution pattern of contact duration from real-world taxi traces, the authors in [89] discussed empirically that the behavior of the contact duration below the characteristic time can be attributed to the traffic control and management aspects, whereas the behavior of the contact duration above characteristic time can be attributed to the vehicular density. Additionally, Ho *et al.* demonstrated that the traffic controls and stops can lengthen the communication links between the vehicles which can delay the characteristic time.

Lastly, there exist a numerous works in the literature that has demonstrated that the mobility pattern of the vehicles has a profound influence on the spatio-temporal dynamics of vehicular connectivity, specifically V2V connectivity. However, the lack of emphasis in understanding the nature of V2I connectivity resulted in understanding the V2I connectivity pattern, specifically in PTNs, as will be discussed in Chapter 6.

2.3.1 Robustness Analysis

Robustness analysis is a topic of hot research area in the field of network theory. Following the work published by Albert [92], the domain attracted a significant

research interest. Many real-world complex systems have demonstrated an exceptional ability to retain their basic functionality even in the case of failure of their network components, termed as the network robustness. It has been illustrated in a numerous works in the literature that the network topology plays a vital role in assessing the network robustness [92, 93, 33]. Percolation theory, which is one of the subfields of statistical physics and mathematics, is typically employed in assessing the condition for a network breakdown under various node removal (random and targeted), and attack models (random dynamic, random static, targeted dynamic, etc.).

The primary interest in robustness analysis is the impact of node failures on the integrity of the network. By randomly removing a fraction of nodes (f), a limited impact is observed on the network's overall integrity. However, increasing f can isolate chunks of nodes from the giant component, and finally, for a sufficiently large f, the giant component breaks into tiny disconnected components [5, Chapter 8]. Thus, to understand the consequences of node removal accurately, understanding the underpinnings of inverse of the percolation process is a pre-requite.

Barabási [5] in his work has provided the detailed mathematical derivations regarding the critical threshold (f_c) required to significantly effect the network's integrity for both random- and scale-free network topologies, and interested reader is referred to the same. There exist a numerous works in the literature where the authors have demonstrated that the scale-free network topology can withstand an arbitrary level of random failures without significantly effecting the network's integrity, in which the hubs are responsible for this remarkable robustness. Whereas, random network topology is more susceptible to random failures which readily fragments into isolated subgraphs at the critical threshold [5, 94, 95, 96, 97].

In our work, we evaluate the robustness of the vehicular network connectivity using the approach of random node removal to demonstrate the capacity of the network to sustain malfunctions, and yet continue to offer the best possible service, as will be discussed in Chapter 6.

2.4 Summary

The application of network theory has permitted in-depth understanding of the connectivity, topological behavior and operations of many practical networked systems as well as the roles the 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 interested us to work in the field of public transport analysis from a graph theoretic perspective.

Furthermore, contemplating the multiple transport networks as independent and non-interacting networks in a numerous PTN analyses accomplished in the literature has led to evasive results considering the real-world usage of the network. Thus, in view of the practical relevance of quantifying the actual network behavior, the need for modeling the interdependency of multiple transport modes is emphasized in our work which had attracted limited contribution in the past.

Urban vehicular networks are expected to be an integral part of future intelligent transport system to mitigate serious on-road issues where the vehicles with on-board wireless communication devices can exchange information. A numerous works in the literature had demonstrated that the mobility pattern of the vehicles had a profound influence on the spatio-temporal dynamics of vehicular connectivity, specifically V2V connectivity. However, the lack of emphasis in understanding the nature of V2I connectivity, specifically in PTNs, and the never explored propensity of analyzing the impact of a PTN topology on the nature of vehicular network connectivity interested us to work in the direction. Though the afore mentioned topics are extensively explored in the areas of traffic and transport engineering fields, the analysis from a graph theoretic perspective demands some serious attention. Lastly, in its first kind, the robustness of the vehicular network is verified using the approach of random node removal to demonstrate the capacity of the network to sustain malfunctions, and yet continue to offer the best possible service.

Chapter 3

Representation of Public Transport Network as A Graph

A public transport network can be either unimodal or multi-modal in nature. In our work, we emphasize on the topological analysis of two different modes of public transportation, namely, the bus transport network (BTN), and the metro transport network (MTN), since the two modes are widely used by the public to meet their daily commuting needs [98]. The current chapter focuses on the topological representation of a public transport network using the graph theory concepts. The major contributions of the chapter are three fold:

1. In Section 3.1, a step-by-step procedure is discussed to generate a public transport network topology from the real-world datasets by applying the graph theory concepts.

2. Unlike the conventional graph representation employed in the previous works, in Section 3.2, we propose a novel approach called *supernode graph representation* to model the public transport network topology.

3. In Section 3.3, we generate a weighted graph structure by adding a weight to both a node and an edge. We propose a *static demand estimation* approach to add a node weight, and the load between the nodes is made use of as an edge weight.

Based on the various statistical information like, the network size, number of daily passengers benefiting from the transport modes, and contribution of the two chosen transport modes to the overall public transportation in a city, we

consider three cities, namely, Hong Kong, London, and Bengaluru in our studies. Though the main focus of the current chapter is on the topological representation of bus transport network, in general, it is applicable to both the bus and metro transport networks.

3.1 Data Analysis

Although significant research interest in the field of network theory has been cultivated for several decades, the established concepts have been applied to the real-world data only in the recent past, mainly due to the availability of real-world datasets and the high-end tools to process such huge datasets. With the aid of real-world datasets, a network topology which closely mimics the real-world structure can be generated using the concepts of graph theory. In this section, we discuss the detailed procedure in building a PTN topology using the real-world datasets which involves three steps: data collection, data mining and data visualization.

3.1.1 Data Collection

In this section, the various means for extracting real-world data are discussed along with the information encased in the datasets. To the best of our knowledge, there is no single source to extract the dataset on various public transport networks, however, by referring to the literature, and based on our observation, a list of online sources and relevant datasets are given in Appendix B. The extracted datasets include information on

1. List of *stops/stations* along with their id's, names, latitude/northing and longitude/easting data.

2. 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 the 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,

ROUTE_ID	COMPANY_(	ROUTE_NAN	SERVICE_MC	SPECIAL_TYF	LOC_START_	LOC_END_NAI	FULL_FARE	LAST_UPDAT
1001	KMB	1	R	0	Chuk Yuen Esta	Star Ferry	$5.80	9/24/2015
1002	KMB	10	R	0	Choi Wan	Tai Kok Tsui (Circ	$4.90	4/19/2012
1006	KMB+CTB	102P	T	1	Shau Kei Wan	Mei Foo	$9.80	4/19/2012
1008	KMB+CTB	103	R	0	Chuk Yuen Esta	Pokfield Road	$9.80	4/19/2012
1009	KMB+CTB	103P	T	1	Pokfield Road	Mong Kok (Yim F	$9.80	10/19/2012
1010	KMB+NWFB	106	R	0	Wong Tai Sin	Siu Sai Wan (Isla	$9.80	4/19/2012
1014	KMB	108	R	0	KAI YIP	Braemar Hill	$10.80	4/19/2012
1015	KMB+NWFB	109	R	0	Ho Man Tin	Central (Macau F	$9.30	10/15/2014
1016	KMB	11	R	0	Diamond Hill S	Kowloon Station	$5.10	4/19/2012
1017	KMB+NWFB	110	R	0	Shau Kei Wan	Tsim Sha Tsui (M	$9.30	4/19/2012
1019	KMB+NWFB	112	R	0	North Point	So Uk	$9.30	10/19/2012
1021	KMB+NWFB	115	R	0	Kowloon City F	Central (Macau F	$9.30	4/19/2012
1022	KMB+NWFB	115P	T	1	LAGUNA VERDI	Central (Macau F	$9.30	4/19/2012
1023	KMB+NWFB	116	R	0	Tsz Wan Shan (	Quarry Bay	$9.80	4/19/2012
1024	KMB+CTB	117	T	1	Sham Shui Po (	Happy Valley (Lo	$9.30	4/19/2012
1027	KMB	11B	R	0	Kwun Tong (Ts	Kowloon City Fe	$5.10	6/30/2013
1028	KMB	11C	R	0	Chuk Yuen Esta	Sau Mau Ping (U	$5.80	4/19/2012
1029	KMB	11D	R	0	Lok Fu	Kwun Tong Ferry	$5.10	4/19/2012
1030	KMB	11K	R	0	Chuk Yuen Esta	Hung Hom Static	$5.10	4/19/2012
1031	KMB	11X	R	0	Sau Mau Ping (	Hung Hom Static	$7.50	4/19/2012

(a)

ROUTE_ID	ROUTE_SEQ	STOP_SEQ	STOP_ID
1001	1	1	4001
1001	1	2	4002
1001	1	3	4003
1001	1	4	4004
1001	1	5	4005
1001	1	6	4006
1001	1	7	8090
1001	1	8	4008
1001	1	9	4009
1001	1	10	4025
1001	2	1	4025
1001	2	2	965
1001	2	3	4076
1001	2	4	4027
1001	2	5	9009
1001	2	6	193
1001	2	7	9053
1001	2	8	751
1001	2	9	4032
1001	2	10	4001

(b)

STOP_ID	STOP_TYPE	X	Y	LAST_UPDAT
2	1	843622	813951	4/19/2012
3	1	843333	814109	4/19/2012
4	1	842947	814013	4/19/2012
5	1	842646	813814	4/19/2012
6	1	842419	813736	4/19/2012
7	1	842307	813674	4/19/2012
8	1	842182	813741	4/19/2012
9	1	842149	813996	4/19/2012
10	1	842092	814350	4/19/2012
11	1	842111	814519	4/19/2012
12	1	842016	814846	4/19/2012
13	1	841835	815086	4/19/2012
15	1	836740	815798	4/19/2012
16	1	836554	815722	4/19/2012
17	1	831457	840927	4/19/2012
18	1	836068	815612	4/19/2012
19	1	835644	815588	4/19/2012

(c)

Fig. 3.1: An excerpt from the Hong Kong BTN dataset. (a) The columns from left to right: list of bus routes, name of the operator , route id, service mode (R: Regular, T: Regular and specific time, N: Night etc.), special type (0: all time, 1: time or day specific), start location, end location, end-to-end travel fare, and last updated information. (b) The columns from left to right: route id, route sequence (1: outbound, 2: inbound), stop sequence along the route id 1001, and their stop id's, i.e., route No. 1001 has 10 stops along the outbound journey route with 4001 being the first and 4025 being the last stop along the route, and vice versa for the inbound journey; (c) The columns from left to right: the list of bus stops along with their spatial information about easting (x) and northing (y).

peak hours, off-peak hours, day-time, night-time routes), etc., are also available in a few datasets. Fig. 3.1 shows an excerpt from the dataset of Hong Kong bus transport network.

3.1.2 Data Mining

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 as compared to gigantic social networks comprising of billions of nodes and edges which are computationally expensive. Here, we describe three basic steps to extract meaningful information by mining the crude datasets:

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.

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

1. Since PTNs belong to the category of spatial networks, understanding the topological behavior along with spatial information would facilitate better network analysis. The spatial information of public transport stops listed in the datasets are either easting-northing or latitude-longitude. Since many of the network visualization tools adopt 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 [99]. 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 [100].

2. In some datasets, the numbers assigned for the stops are typically non-

sequential in nature which poses computational challenges, e.g., in generating adjacency matrix. 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.

3. The concept of short distance station pairs (SSPs) has been commonly adopted to represent a group of stations as a single (merged) station [17, 18, 22]. Assigning new id's to such SSPs according to the sequential mapping carried out in Step 2 is recommended to facilitate easy 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 a few works [17, 22]. The idea behind identifying SSPs is to establish a virtual connectivity among the nodes. While 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_{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_{th} creates a lot of SSPs in a dense network, whereas a large value of d_{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_{th} may bias the understanding of network behavior. Hence, a careful selection of the d_{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} is 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 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 discussed in Section 2.1.1 in Chapter 2. 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.

3.1.3 Data Visualization

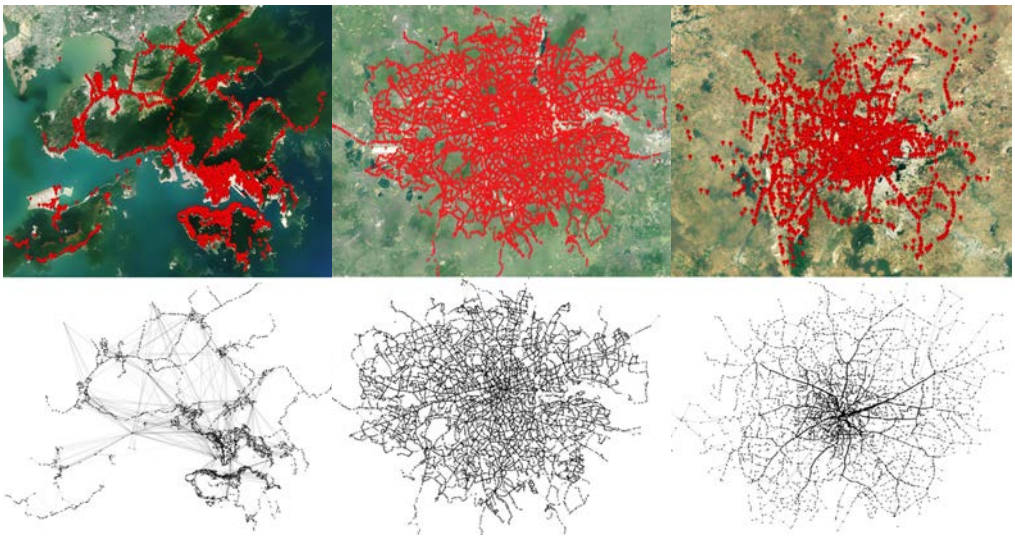
For visualizing a network, there are many open source network visualization tools, and the selection would depend on the need of the analysis. For a comparison of different visualization tools, interested reader is referred to ref. [101].

Thus, in the process of building a real-world network topology, we suggest three fundamental steps, i.e., data collection from a reliable source, data mining in order to reduce the anomalies, make computation less arduous, define the network topology based on the graph type and the space-of-representation, and lastly, data visualization to perceive the network topology. In our work, for the topological analysis of the three cities, we employ a *directed* and *weighted* graph structure represented in *L-space*.

3.2 Supernode Graph Representation

In this section, we first discuss the representation of a public transport network topology as a conventional graph, which is later modified with the proposed supernode graph representation. Using the conventional graph theory approach, a graph G is defined by (2.1) and (2.2) as discussed in Section 2.1 in Chapter 2.

Considering the spatial embedding of a public transport network, in our work,



(a) **(b)** **(c)**

Fig. 3.2: Spatial location of bus stops for (a) Hong Kong; (b) London; and (c) Bengaluru networks.

a graph G is represented by $G = (V(x, y), E)$ where V and E are described as:

$$V(x, y) = \{n_i\,(x_i, y_i) \mid x_i = \text{latitude}, y_i = \text{longitude};\ \forall i = 1, 2, ..., N\} \tag{3.1}$$

$$E = \{e_{ij} \rightarrow (n_i\,(x_i, y_i)\,, n_j\,(x_j, y_j))\,\forall\, n_i(x_i, y_i) \in V;\ \forall i = j = 1, 2, ..., N; i \neq j\} \tag{3.2}$$

In the subsequent chapters, $n_i(x_i,\ y_i)$ is represented as n_i assuming that a particular node is always identified by its latitude and longitude. Fig. 3.2 shows the spatial location of the bus stops, and the corresponding bus transport network structure for the three chosen cities represented as a conventional graph.

Additionally, the inspection of spatial embedding of nodes in bus transport networks resulted in a new network element called *supernode*, as shown in Fig. 3.3. A supernode is a set of geographically closer nodes which satisfy the condition:

$$d_{ij} \leq d_{th}\ ;\ \ d_{th} = 100\text{m} \tag{3.3}$$

where d_{ij} is the geographic distance between two nodes i and j, and d_{th} is a predefined threshold distance. The value of d_{th} is set to be 100 m in our work assuming it to be an optimal walkable distance to reach a nearby station. If $d_{ij} \leq d_{th}$, nodes i and j are said to be geographically close to each other, and are combined to represent a single node called supernode. The geographical distance between the nodes (d_{ij}) is evaluated using the Haversine formula [102]. The

combining of multiple nodes as a supernode is not a physical phenomenon, and instead is merely a structural reorganization.

Fig. 3.3: An example of geographically closer nodes in Hong Kong.

Thus, a supernode graph structure consists of four components: regular node set (V), supernode set (V_S), regular edge set (E), and superedge set (E_S), i.e., $G = (V(x,y), V_S(x,y), E, E_S)$, where V_S and E_S are given by (3.4) and (3.5) respectively.

$$V_S = \{sn_i\} \ \forall i = 1, 2, .., N_S. \tag{3.4}$$

where $N_S = |V_S|$, V_s is the set of all supernodes in the network where, each supernode is defined as: $sn_i = \{n_j, n_k\}$, such that $d_{jk} \leq d_{th}$, i.e., each supernode is a set of two regular nodes whose geographic distance is less than or equal to d_{th}. If $(sn_i \cap sn_j) \neq \emptyset$, then, $\tilde{sn_i}$=$\{sn_i \cup sn_j\}$, i.e., if there exists some nodes that are common to multiple supernodes, these supernodes are combined to form a giant supernode ($\tilde{sn_i}$), which is assigned with a unique node id for its easy identification. A giant node is also within the supernode set. The newly formed supernode is assigned a new spatial location which is the mean location value of the corresponding elements, i.e., $sn_i(x_i, y_i) = \{n_j(x_j, y_j), n_k(x_k, y_k), ...\}$, where $x_i = \text{mean}(x_j, x_k, ...), y_i = \text{mean}(y_j, y_k, ...)$. Figure 3.4 illustrates an example of how a supernode and a giant supernode are defined. In our analysis, considering the three cities, a supernode comprised of two regular nodes, and a giant supernode comprised of 9 to 12 regular nodes.

Assuming that an edge e_{ij} exists between nodes n_i and n_j, a superedge can

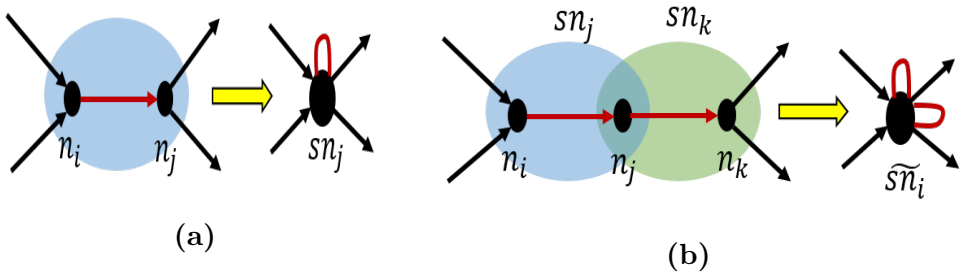


Fig. 3.4: An example of (a) supernode; and (b) giant supernode.

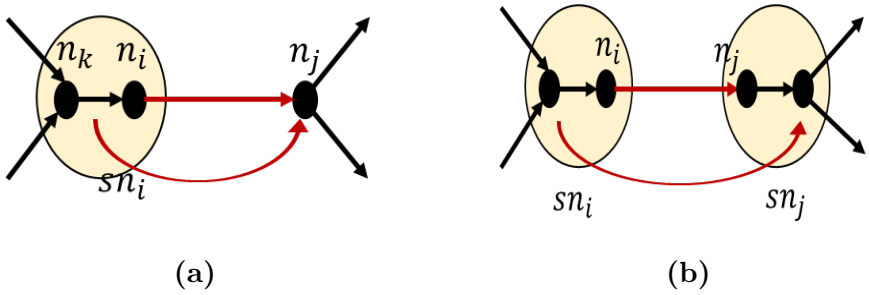


Fig. 3.5: An example of defining edge between (a) a supernode and a regular node; and (b) two supernodes.

then be defined between two supernodes as $e_{sn_i,sn_j} \rightarrow (sn_i, sn_j) \forall (n_i \in sn_i, n_j \in sn_j)$ as shown in Fig. 3.5b, or a superedge can be defined between a supernode and a regular node as $e_{sn_i,n_j} \rightarrow (sn_i, n_j)\ \forall (n_j \in V, n_i \in sn_i)$ as shown in Fig. 3.5a. Thus, the superedge set is defined as

$$E_S = \left\{ e_{sn_i,sn_j} \cup e_{n_i,sn_j} \right\} \forall sn_i, sn_j \in V_S, n_i, n_j \in V \tag{3.5}$$

In the course of defining the supernode structure, some of the original nodes and self loops are eliminated. The self loops convey information regarding the edge weight, however, in our analysis, since we do no use the edge weight (except for path length calculation), the removal of self loops have no impact on the overall analysis. Intuitively, based on our condition for forming a supernode, it can be stated that the existence of supernodes is highly probable in bus networks as compared to metro networks. The idea behind identifying supernodes is to establish a virtual connectivity among the nodes to make the analysis more practical. However, it should be noted that choosing an extremely small value of d_{th} creates numerous supernodes in a dense network, whereas a large value of d_{th} is meaningless, since a long walking distance to reach a nearby station is unreasonable. Hence, a judicious selection of the d_{th} is important.

The significance and the necessity of supernode graph representation in a PTN analysis are summarized as follows:

1. Combining the geographically significant nodes in a network benefits in improved understanding of the network behavior. For example, combining nodes with significant degrees in a network helps identify hubs in a network, which in turn aids in determining the convenient switching/transfer points in a network. Not all the supernodes are considered the preferred transfer points, however, the supernode representation offers a strong foundation in identifying such nodes in the network.

2. Supernode structure helps eliminate redundancies in the network for fast computation (compared to conventional graph representation), and yet retains the original network structure without significant loss, i.e., with the supernode structure, we can reproduce a network that nearly mimics the original network with a reduced dataset.

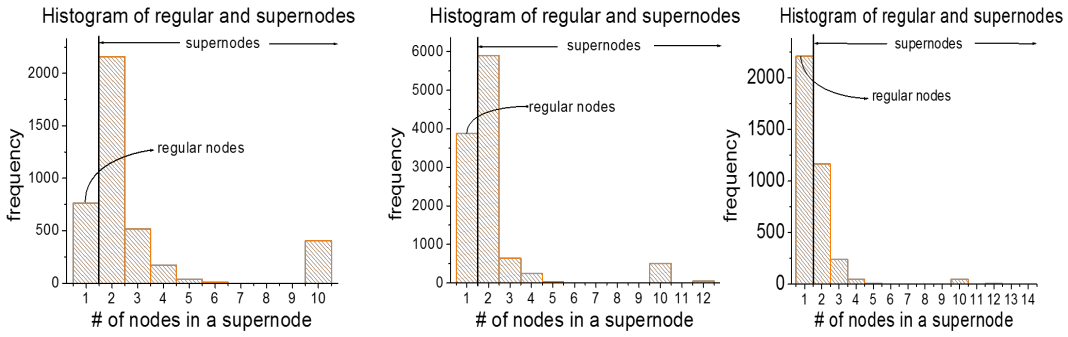
   A parameter *redundancy* is defined to examine the level of reduction between the original network and the supernode network which is defined as:

$$Redundancy = \frac{\#\text{ of nodes with supernode graph representation}}{\#\text{ of nodes with conventional graph representation}} \quad (3.6)$$

   Table 3.1 gives statistical details of the original network, the supernode network, and the average redundancy observed for the three networks. Figure 3.6 indicates the histogram of the regular and supernode counts for the three cities. To define the redundancy, we consider the smallest available division of the land area such as a district or a ward as shown in Fig 3.7, and we term it as a *zone*. The idea behind considering the zones in a network is to observe the redundancy pattern at local level, i.e., observing the reduction of nodes within a chosen zone considering the conventional and supernode graph representation. Interestingly, observing the redundancy distribution pattern reveals intriguing information on the distribution pattern of nodes (bus stops) itself. For example, a high density of nodes in certain zones yields more supernodes in the zone causing maximum redundancy, since, high density of nodes indicate more closely packed nodes, and in the

Tab. 3.1: Comparison of network size for the three cities with original and supernode representations.

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



(a) **(b)** **(c)**

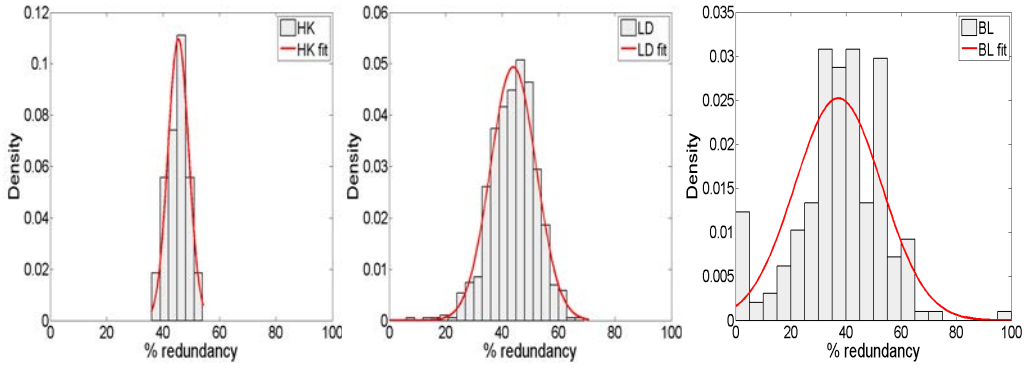
Fig. 3.6: The histogram indicating the number of regular nodes and supernodes in (a) Hong Kong; (b) London; and (c) Bengaluru cities.

supernode representation, such close associates are merged together to represent a supernode. Empirically, the redundancy distribution is observed to follow a Normal distribution for the Hong Kong, London and Bengaluru networks with μ equal to 45, 44, 37, and σ equal to 4, 8, 16 respectively as shown in Fig. 3.8. It can be inferred from both the Table 3.1, and Fig. 3.8 that, the redundancy is maximum in the case of Hong Kong bus network as compared to the London and Bengaluru networks, indicating that the bus stops in Hong Kong city are geographically closer and denser in a given zone. Thus, by observing the redundancy pattern in a network, a better insight into the node distribution pattern can be procured.

Also, as tabulated in Table 3.1, in all the three networks, an average of

(a) **(b)** **(c)**

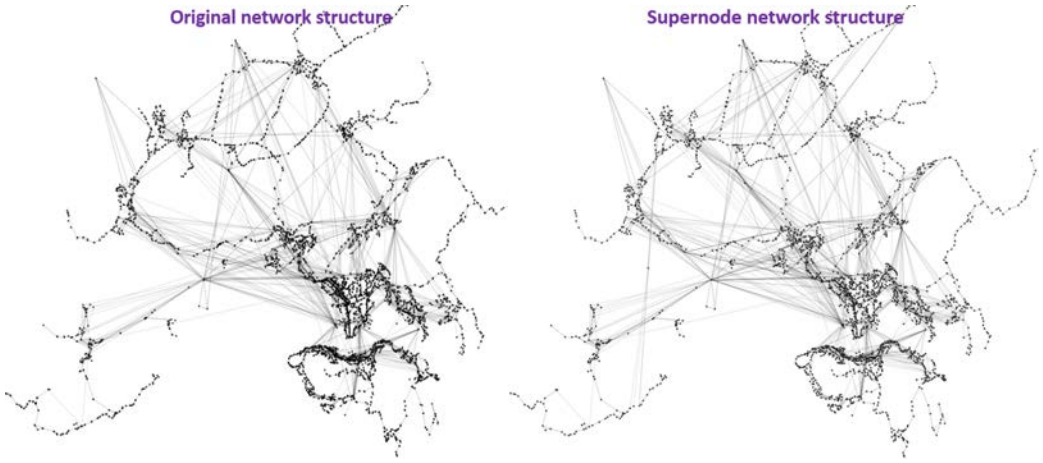
Fig. 3.7: The consideration of smallest available geographical division of a land area as a zone in the (a) Hong Kong; (b) London; and (c) Bengaluru cities. For the Hong Kong city, a zone represents a district, however, for London and Bengaluru cities, a zone represents a ward. The difference in considering the geographical area as zone is primarily based on the availability of real-world data from the respective cities.



(a) **(b)** **(c)**

Fig. 3.8: Redundancy distribution for (a) Hong Kong; (b) London; and (c) Bangalore cities.

40% redundancy of nodes is observed, whereas, only 20% redundancy is observed with respect to the edges (network connectivity), indicating that a major portion of the original network structure is still retained even after the network is reorganized as a supernode structure as shown in Fig 3.9. The 40% reduction in nodes actually reduces redundant information since the supernodes indirectly hold the information regarding the geographically closer nodes in the network, thus retaining the basic information. On the



(a) **(b)**

Fig. 3.9: Hong Kong bus transport network structure in (a) conventional; and (b) supernode graph representation types.

other hand, the 20% reduction in the edges is due to the elimination of self loops which has less influence in our analysis since the information conveyed by the self loops is not used in our analysis.

Thus, we claim that the supernode network representation closely mimics the original network with a reduced dataset, and the information loss in the new network representation is very minimal thus contributing towards capturing the actual network behavior. As we demonstrate in Chapter 4, supernodes play vital role in capturing the actual network behavior which the conventional representation fails to capture.

Therefore, in the analysis of a PTN structure, a supernode graph representation can be practiced in the future for faster computation and better understanding of the network behavior. In Chapter 4, we empirically show that the supernode approach aids in a more practical and improved understanding of the network behavior as compared to the conventional graph representation.

3.3 Weighted Graph Representation

In this section, the supernode graph structure generated in Section 3.2 is modeled as a weighted graph by adding a weight to both the node and edge. In classical graph theory, assigning a weight to an edge is of a common practice to generate a weighted network, however, assigning a weight to a node is seldom considered

[27, 19]. In our work, we choose to assign a weight to both a node and an edge since the added weights assists in a more realistic network analysis as discussed in the following sections.

3.3.1 Node Weight

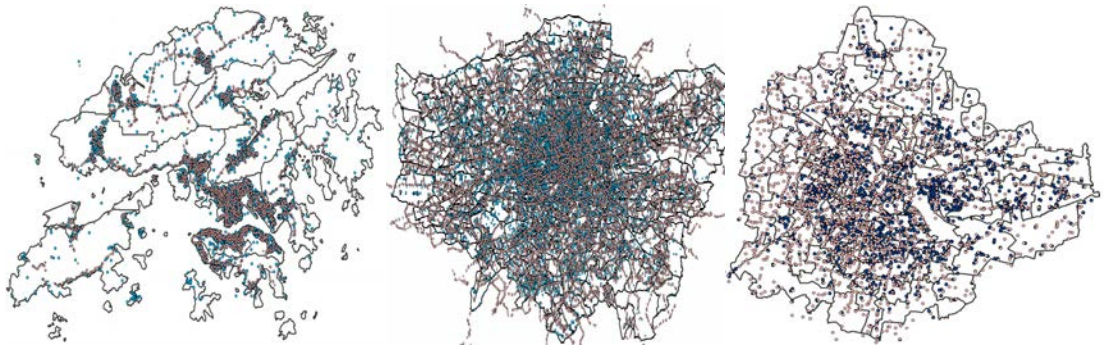
In the literature, the number of bus routes servicing a bus stop (degree) is typically used as a node weight by some authors [18, 27, 28]. However, in our work, *a static demand estimation* approach is proposed to weigh a node's significance. The crux of the static demand estimation approach is that, the utilization factor of a bus stop in a city is greatly influenced by the presence of points-of-interests (POIs) around the bus stop, where a POI can be either be a hotel, office, school, hospital, sports arena, cinema hall, shopping complex or residential apartment. Additionally, a new metric termed *node usage probability* (NOP) is proposed which is defined as the total number of people accessing a bus stop. We believe that, by studying the geographic locations of POIs around a bus stop alongside its node occupying probability, a better idea on the static demand serviced by a bus stop can be procured. The details of the demand estimation approach are as discussed below:

1. The POIs in a city is broadly divided into four categories (m), as shown in Table 3.2. Each category can be assigned a weight to indicate the extent to which the POI attracts the passengers which aids in effortlessly marking the nodes serving maximum demand in a network based on that specific POI category.

2. In order to evaluate the node occupying probability (total number of people accessing a particular bus stop), a small local area (typically a few meters of walkable catchment) around a bus stop is a requisite. However, for the purpose of empirical analysis, and due to the lack of real-world data, in our work, a geographical area termed *zone* is employed as discussed in Section 3.2, such that each zone possesses a defined land area, population, and POIs within it. Fig. 3.10 shows the geographic location of the bus stops, and the POIs located in various zones for the three cities.

Tab. 3.2: Different POI categories considered for static demand estimation method.

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



(a) **(b)** **(c)**

Fig. 3.10: Spatial location of the bus stops (pink color), and the POIs (blue color) for (a) Hong Kong; (b) London; and (c) Bengaluru cities.

Thus, based on the statistical information on POIs around a node, and its NOP, we assign a weight to a node as below:

$$(w_i)_{\text{zone}} = \left(\frac{c_1 \sum_{m=1}^{4} \rho_m + c_2 \rho_p}{\rho_N} \right)_{\text{zone}} + c_3 k_i \quad ; \text{zone} = 1, 2, ..., z \tag{3.7}$$

where $(w_i)_{zone}$ is the weight of a node i in a specific zone. The term ρ_m denotes the POI density of category m in a zone, ρ_p denoted the static population density in a zone, ρ_N is the bus stop density in a given zone, k_i is the node degree (as discussed in Section 2.1.2 in Chapter 2), the fraction $(\rho_p/\rho_N)_{\text{zone}}$ denotes the node occupying probability in a given zone, and $(\rho_m/\rho_N)_{\text{zone}}$ indicates the number of bus stops servicing the number of POIs in a zone. For simplicity, the constants c_1, c_2 and c_3 are assumed to be equal to 1 (for the dynamic demand estimation, fine tuning of the three constants play a vital role). To ensure data integrity across

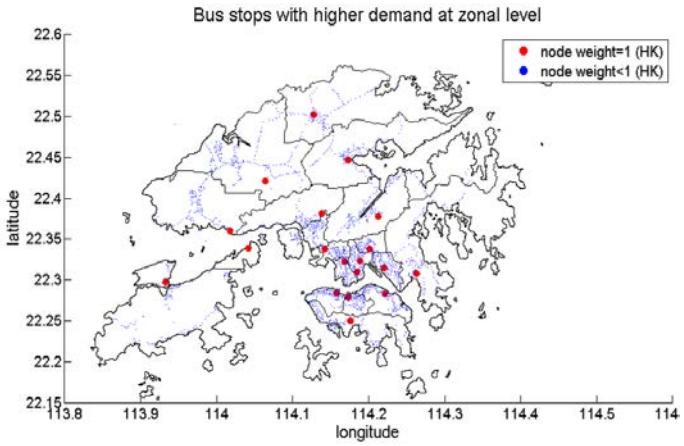
all the zones, and for the ease of representation, the node weights are normalized at zonal level, i.e.,

$$(w_{i\_\text{norm}})_{\text{zone}} = \left(\frac{w_i - w_{\min}}{w_{\max} - w_{\min}} \right)_{\text{zone}} \quad ; \text{zone} = 1, 2, ..., z \tag{3.8}$$

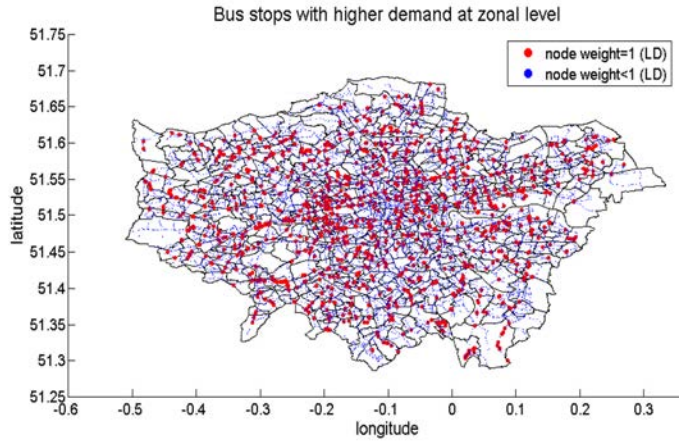
In equation (3.8), closer the value of w_{i_norm} to one, higher the demand serviced by the node in the chosen zone. Since the normalization is carried out at zonal level, each zone will have at least one node with w_{i_norm} equal to 1 indicating the node(s) servicing higher demand at zonal level. Fig. 3.11 indicates the nodes (in red color) serving maximum demand in the three cities at zonal level based on the normalized node weight calculated using (3.8). Table 3.3 demonstrates the approach used to assign node weight to a few nodes in a specific zone in the Hong Kong city.

Unlike in a few works in the literature where the node degree was solely employed to quantify a node's significance [18, 27, 28], in the static demand estimation approach, the degree combined with local demand is taken into consideration to quantify a node's significance. An advantage of the approach is that, the bus stops that attracts significant crowd at zonal level can be better understood. With the evolution of the city, these zonal significant nodes tend to behave as the future hubs offering more service to the people, and may act as future congestion hotspots if timely measures are not taken.

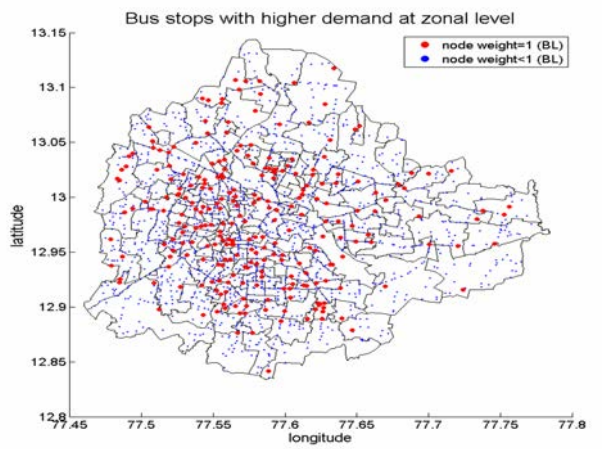
From (3.7), we can notice that, a node weight is a function of four parameters, i.e., ρ_m, ρ_p, ρ_N and k_i. As compared to the other three parameters, ρ_p has a strong influence on w_i since the rate at which population density in a zone increases more drastically as compared to the rate at which other three parameters increase. However, in our approach, since all the parameters in (3.7) are static, the demand in a zone typically follows the node degree as shown in Fig. 3.12 which is plotted from the real-world dataset provided by the KMB (Kowloon Motor Bus Co.) bus operator in Hong Kong. From (3.7), it is also observed that, if the nodes in a particular zone have comparable degree's, then the proposed demand estimation approach fails to effectively capture the local demand serviced by the nodes in that zone, as observed in the case of London and Bengaluru cities in Fig 3.11.



(a)



(b)



(c)

Fig. 3.11: The bus stops servicing higher static demand at zonal level for (a) Hong Kong; (b) London; and (c) Bengaluru cities.

Tab. 3.3: Illustration of the node weight approach to assign the a node weight to a node (i) in a specific zone (Central and Western district) in Hong Kong city.

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

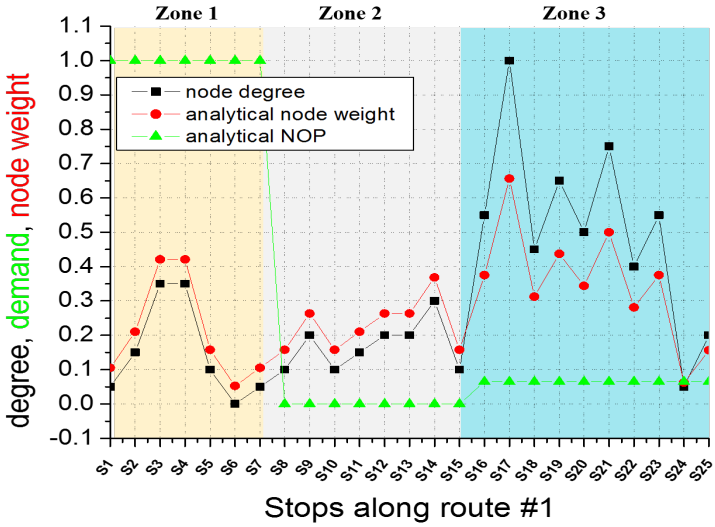


Fig. 3.12: An example demonstrating the dependency of a node weight on it's degree for the bus stops along route #1 in Hong Kong city. The stops S1:S7 belongs to zone 1, S8:S15 belongs to zone 2, and S16:S25 belongs to zone 3.

Tab. 3.4: Classification of zones into emitter/absorber/neutral regions.

# jobs	# working population	# POIs	Zone type
scarce	scarce	scarce	Neutral
scarce	scarce	abundant	Absorber
scarce	abundant	scarce	Emitter
scarce	abundant	abundant	Emitter and Absorber
abundant	scarce	scarce	Absorber
abundant	scarce	abundant	Absorber
abundant	abundant	scarce	Emitter and Absorber
abundant	abundant	abundant	Absorber

Zone Classification Approach

To gain better insight on a node's behavior, in this section, we assess the behavior of a zone together with the static demand estimation approach. That is, we consider three information category types: # of POIs, # of working population, and # of jobs in a specific zone to classify a zone as either an *emitter*, *absorber* or *neutral* type as shown in Table 3.4. The details regarding the three information categories at zonal level are extracted from [103, 104, 105]. A category type in a particular zone is considered *scarce* if its value is less than the mean or median value of the respective distribution, otherwise, it is considered *abundant*. The emitter zone results in out-flux of passenger from the zone, an absorber zone results in in-flux of passengers from neighboring emitter zones, and neutral zone results in a balanced in-flux and out-flux of passengers.

The intent of knowing a zone's behavior is that, zone classification approach together with the node weight approach assists in better understanding a node's behavior. That is, a highly significant node(s) within an emitter zone is expected to serve maximum out-flowing demand in the zone, and a highly significant node(s) within an absorber zone serves the maximum in-flowing demand. Figs. 3.13a - 3.13c shows the behavior of each zone as either an emitter, absorber or neutral type. From the real-world data samples obtained from KMB operator in Hong Kong, we compare the empirical demand using the proposed node weight

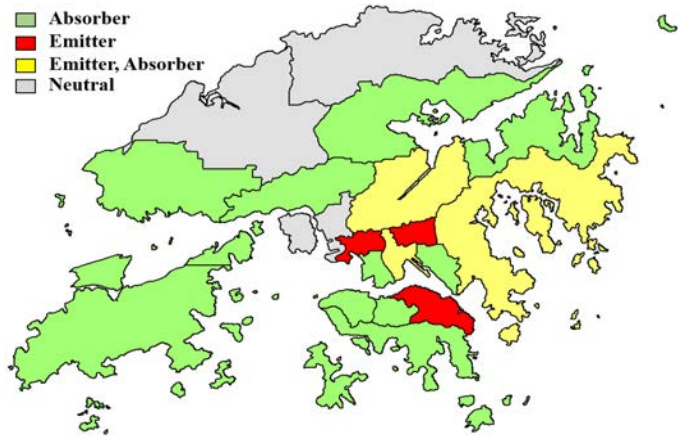
approach with the real-world demand, and the results are as shown in Fig. 3.14b. Fig. 3.14a indicates the GPS traces for the route #1 alongside with the bus stops, POIs, and zones. Based on the zone classification approach, the bus stops within zone 1 (S1-S7) acts as emitter stops, bus stops within zone 2 (S8-S15) and zone 3 (S16-S25) acts as either emitter or absorber stops. Combining the approaches of zone classification and node weight, we can state that, a bus stop with high demand within an emitter zone results in an increased boarding of passengers, and a bus stop with high demand within an absorber zone results in an increased alighting of the passengers.

Accordingly, from the real-world data samples, the stops S1, S9, S16, S18, S21, and S23 are regarded high demand nodes during morning rush hours, the stops S1, S9, S16, S17, S18, S21, S23 are regarded high demand nodes during evening rush hours. Comparing the results with our proposed approach, the stops S3, S4, S9, S14, S16, S17, S18, S21, S23 are regarded high demand nodes which leads to a 67% accurate prediction of the actual demand. Thus, the proposed node weight approach together with the zone classification approach can be a great source of information to the network operators to analyze a node's behavior in the network. Though the 67% accurate prediction is not a better figure, the prediction level is merely 48% without considering the supernode graph representation. Also, a few reasons that affect the prediction accuracy are discussed below.

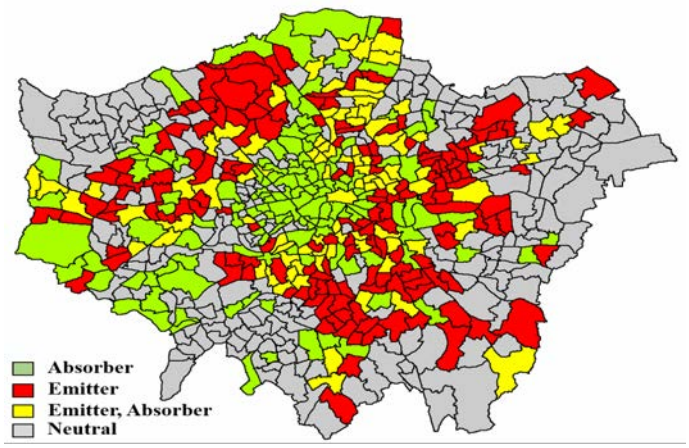
Though the proposed static demand estimation approach aids in identifying zonal significant nodes, the approach has a few constraints due to which it might fail to effectively capture the actual demand, i.e.,

1. In (3.7), the parameters POI density and node occupying probability are not two independent factors, instead, are interdependent. For example, the number of people accessing a bus stop in a zone is dependent on the existence of POI categories like residential apartments, working offices, and schools in that specific zone. Therefore, a more accurate modeling of passenger movement considering the dependency of the two parameters can lead to a better demand estimation.

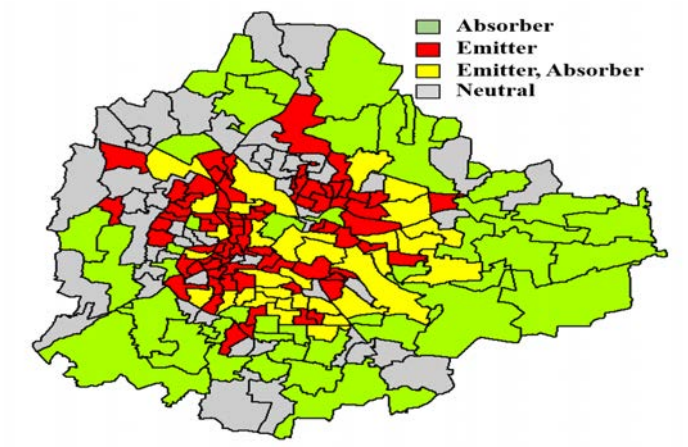
2. The area considered to represent a *zone* is far from a realistic consideration which contributes towards less accurate NOP estimation for a bus stop.



(a)

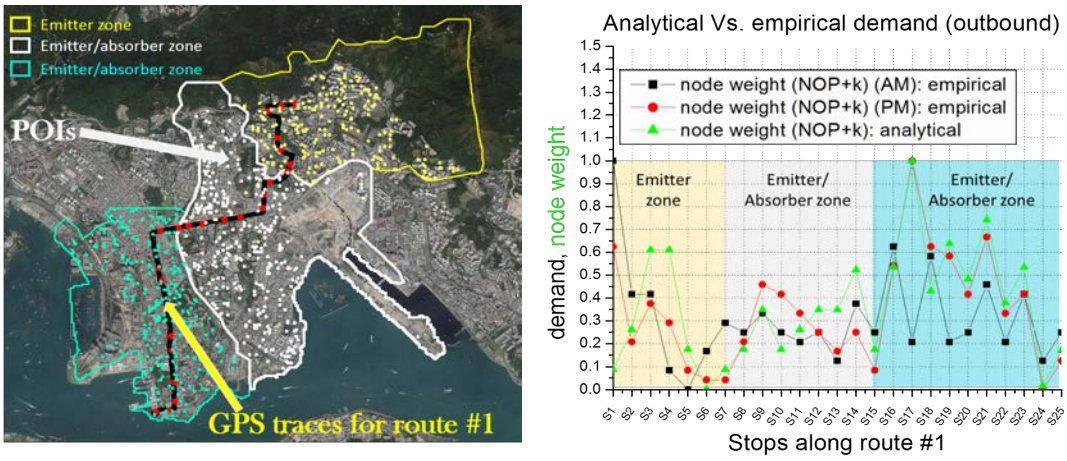


(b)



(c)

Fig. 3.13: Zone behavior as emitter, absorber or neutral type for (a) Hong Kong; (b) London; and (c) Bengaluru cities.



(a) **(b)**

Fig. 3.14: (a) The GPS traces for route #1 in Hong Kong alongside with the bus stops, POIs, and zones; and (b) comparison of analytical and empirical demand serviced by the stops along bus route #1.

> Though, the area chosen is based on the availability of empirical data, a better estimation requires the consideration of an area typically a few meters around the bus stop to determine a better node occupying probability.

Despite there exists a few constraints, the static demand estimation approach alongside with the zone classification approach assists in determining the nodes servicing maximum demand at zonal level, and to understand their behavior as either emitter, absorber or neutral type. The node weight estimation discussed in the current section is primarily for mono-layer analysis, i.e., only for bus transport network. However, an extension of (3.7) considering multiple transport modes is discussed in detail in Chapter 5.

3.3.2 Edge Weight

Similar to a few works in the literature [17, 28, 21], in our work, we choose the number of routes operating between two given nodes as the edge weight, i.e.,

$$w_{ij} = \sum_{i \neq j} r_{ij} \tag{3.9}$$

where r_{ij} is the route operating between nodes i and j. It is observed that, the edge weight distribution for all the three cities follows an exponential distribution as shown in Fig. 3.15a. The long tail end of the distribution indicates the edges

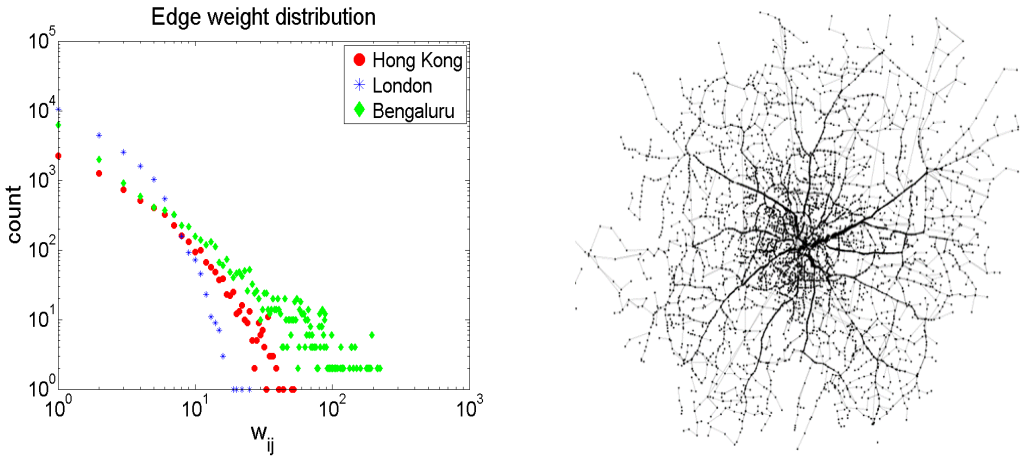


Fig. 3.15: (a) Edge weight distribution for the three cities; and (b) an edge weighted Bengaluru bus network topology.

with maximum weights, and thus the most overlapped routes in the network. From a network analysis perspective it is a great source of information to the operators since these overlapped routes denote the fact that a few sections of the network are being extensively used leading to high traffic congestion during peak travel hours. These edges also act as starting points for the traffic aggregation leading to slow moving traffic, and may even cause longer waiting time for passengers to board the buses due to multiple buses waiting near the bus stops. Such an analysis acts as a major source of information for the applied fields of studies such as route planning and optimization. Fig. 3.15b shows the Bengaluru bus network topology with added edge weights where the maximum edge weights are observed at the city's center. The usage of edge weight has a limited application in our work as compared to the node weight. The edge weight would be only employed in the evaluation of path length as will be discussed in Section 4.4 in Chapter 4.

3.4 Summary

The current chapter emphasized on applying the fundamental concepts of graph theory to represent the topology of a public transport network for a chosen graph type and space-of-representation. A step-by-step detailed procedure was discussed to represent a PTN topology from the real-world datasets that involved

data collection, data mining, and data visualization. Considering the spatial embedding of a PTN structure, the supernode graph structure representation and its significance in modeling the PTN structure were discussed. A static demand estimation approach alongside with the zone classification method was proposed based on various information categories to weigh a node's significance and its behavior as emitter, absorber or neutral type.

To the best of our knowledge, our contribution in generating a weighted graph by assigning a node weight considering the real-world usage of the network is of first in its kind. The discussions in the current chapter plays a fundamental, yet vital role in actualizing the study on public transport network analysis in the later chapters since the course of a defined topology has a profound influence in perceiving both the local and global behaviors of a network.

Chapter 4

Mono-layer Analysis

The current chapter focuses on the topological analysis of the three bus transport networks using the concepts of graph theory. The term mono-layer indicates the consideration of single transport mode in the current chapter, i.e., bus transport, which is extended to multi-layer analysis considering the other transport modes as will be discussed in next chapter. The topological analysis is accomplished with the aid of directed and weighted supernode graph structure represented in L-space as discussed in Chapter 3. The major contributions of the chapter are two fold:

1. The statistical analysis of a few network parameters that are useful in the analysis of a BTN topology is accomplished, i.e., local parameters like degree and clustering coefficient, global parameters like the average degree and average path length, and a few centrality measures like betweenness, hub and authority centralities are analyzed. Unique topological behaviors like scale-free and small-world properties are also studied.

2. The topological efficiencies of the three structures are compared to gather useful information. A semi-realistic simulation in SUMO tool alongside with the real-world dataset procured from the KMB operator in Hong Kong is employed to verify the efficiency of a particular route.

In the course of our analysis, empirically we demonstrate at numerous instances the advantage of using the supernode representation in a BTN analysis as compared to the conventional graph representation.

4.1 Connectivity in Bus Transport Networks

The connectivity pattern of a node is evaluated by a network parameter termed degree as discussed in Section 2.1.2 in Chapter 2. The degree of a node signifies the number of edges incident on it, whereas average degree conveys information about average connectivity of nodes considering the overall network.

Table 4.1 shows the average degree of the bus network for the three cities in regular (conventional) and supernode representations. It is observed from Table 4.1 that the average connectivity of the nodes is slightly improved with supernode representation as compared to the conventional graph representation for all the three BTN structures. It is quite obviously due to the combining of nodes in the supernode representation. However, the additional information obtained by knowing the degree in a SN representation is that it conveys more practical information on the actual number of stops that can be reached from a given stop including the walkable catchment, unlike the regular graph topology which merely conveys information on the immediate neighbors of a node. Though such a connectivity is an inherent topological behavior, it is better understood with supernode representation which the conventional graph representation fails to capture. Thus, understanding the degree of a node in L-space aids in a more realistic BTN analysis.

Tab. 4.1: The average degree of the three bus transport network topologies in regular and supernode representations.

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

A few statistical observations regarding node degree are: (a) the in-degree of a node can be either less than, or equal to the out-degree of a node; (b) nodes with high in-degree tend to behave as sink nodes where a numerous routes terminate their journey, and nodes with high out-degree tend to behave as source nodes

where a numerous routes begin their journey; (c) nodes with high overall degree can be treated as hubs in a network; and (d) in a sparse network, the degree of a numerous nodes will be *two* indicating that a stop is merely connected to its neighboring stops. The key point is that such statistical analysis provides information on basic connectivity pattern of a node in the network which acts as prime information to the operators for a better network planning and optimization.

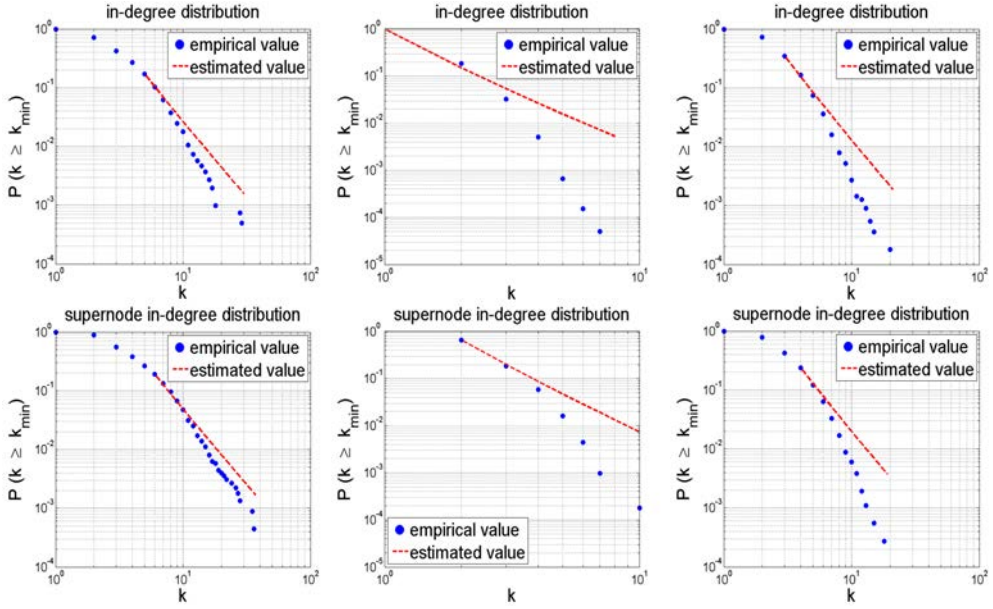
4.2 Are the Bus Transport Networks Scale-free?

Degree distribution indicates the probability of finding a node in the network with a degree k , i.e.,

$$P(k) = \frac{N_k}{N} \tag{4.1}$$

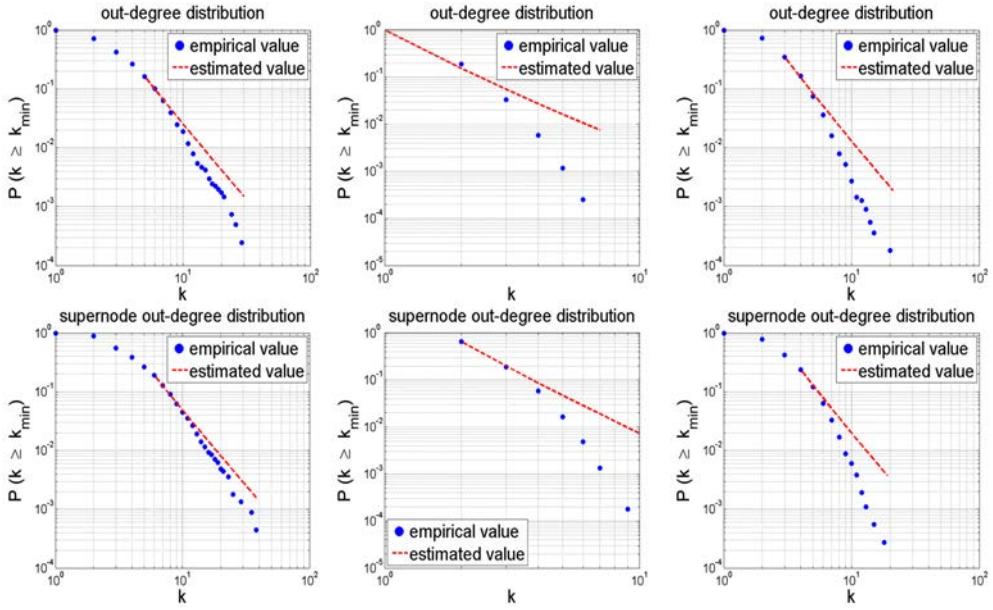
where N_k is the number of nodes with degree k, and N is the total number of nodes in the network. A network whose degree distribution follows a power-law as given by (2.7), is termed as scale-free. For a directed network, the power-law is applied independently to both in- and out- degrees as $P(k_{in}) \propto k_{in}^{-\gamma}$ and $P(k_{out}) \propto k_{out}^{-\gamma}$ respectively. In our work, we follow the approach demonstrated by Clauset *et al.* [106] in fitting the power law to the empirical data. That is, the value of $k_{\min}$ (the lower bound on power law) and γ (the scaling parameter) are evaluated from the empirical data through the maximum-likelihood estimation (MLE), and a Kolmogorov-Smirnov (KS) goodness-of-fit test is employed to evaluate the discrepancy between the empirical and the hypothesized power law distribution, and the corresponding 'p-value' for goodness-of-fit is computed. Typically, if the value of p ≥ 0.1, the power law is a plausible hypothesis.

The graphs in Figs. 4.1 and 4.2 show the power law fit for in- and out- degree distributions for the three bus transport networks in regular and supernode representations. It is interesting to observe that the Hong Kong network fails to satisfy the power law distribution with conventional graph representation, but plausibly follows a power-law with supernode representation (The p-value's for the three cities are tabulated in Table 4.2 with regular and supernode representations). Thus, the inherent scale-free behavior of the Hong Kong bus transport network is better understood with supernode representation, which the conventional representation failed to capture. However, the London and Bengaluru networks show



(a) **(b)** **(c)**

Fig. 4.1: Power law fit for in-degree distribution for (a) Hong Kong; (b) London; and (c) Bengaluru bus transport networks in regular and supernode representations.



(a) **(b)** **(c)**

Fig. 4.2: Power law fit for out-degree distribution for (a) Hong Kong; (b) London; and (c) Bengaluru bus transport networks in regular and supernode representations.

Tab. 4.2: The p-value indicating the goodness-of-fit result between empirical and hypothesized data for the three cities in regular and supernode representations.

	Hong Kong	**London**	**Bengaluru**
Regular	0	0	0
Supernode	0.23	0	0

poor fit to the power-law, and a better fit to the Exponential distribution indicating that both the networks behave as random networks [39, 38]. Thus, a simple modification in the network topology aids in a better and practical understanding of the actual topological behavior, and can assist in exploring the hidden properties of the network. Table 4.3 indicates the values of Exponential (λ) and power-law fitting exponents (k_{min}, γ) for the three bus networks in regular and supernode representations. The exponent values indicated in the table are valid for both the in- and out- degree distributions.

Knowing the degree distribution of a network aids in procuring an obscure idea on the topological evolution of a public transport network in a city, i.e., if the degree distribution follows an exponential law, a newly added node to the network is more likely to be connected with other nodes in a random way, unlike the notion of preferential attachment where the newly added node is connected to the already existing influential nodes in the network.

Also, it is interesting to observe the scale-free property (also called the 80/20 rule) in public transport networks which demonstrate 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, and such networks are free of any scaling applied to the networks. Intuitively, though we presume that there exist certain stops in a network which are serviced by a numerous routes, it is intriguing to verify such a property mathematically. In our work, Hong Kong BTN exhibits such a scale-free property. Considering the actual network behavior in real-world, we assume that the passengers prefer to switch the routes at supernodes in the network, and thus the existence of supernodes plays vital role in determining the scale-free property of a network with respect to its real-world usage.

Tab. 4.3: The Poisson and power-law exponent values for the three bus transport networks in regular and supernode representations.

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

4.3 Cohesiveness in Bus Transport Networks

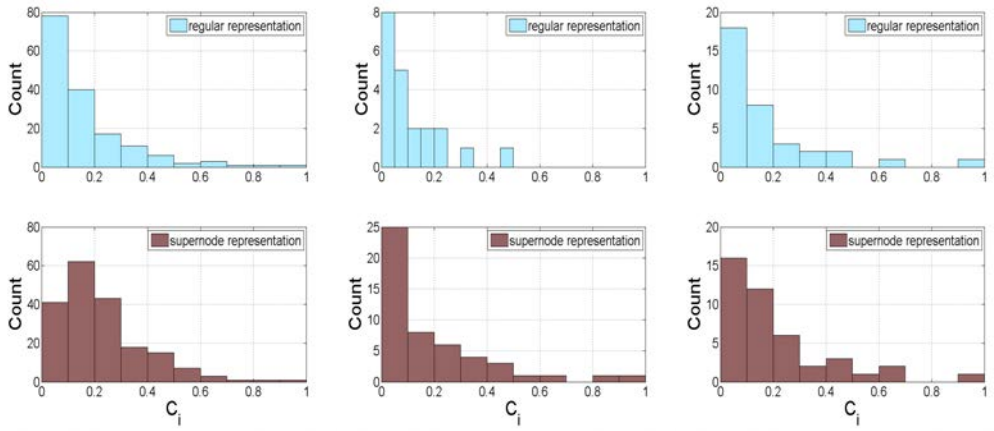
The extent to which the immediate neighbors of a node are connected to each other is examined with the aid of a network parameter termed clustering, which defines the level of cohesiveness in a network. The cohesiveness of nodes is evaluated at local level through local clustering coefficient given by (2.10), and is determined at global level through global clustering coefficient given by (2.11).

Though understanding clustering coefficient by itself has less importance in the analysis of BTN, the relationship between C_i and k raised interesting observations, i.e, as observed from (2.10) and (2.11), the inverse dependency of C_i and k indicates the hierarchical structure of a network in L-space where high degree nodes (hubs) tend to form a numerous connections among its neighbors reducing the possibility of the neighbors to connect among themselves, and thus reducing the local clustering coefficient of high degree nodes, whereas, a low degree node has a greater tendency to be connected among its neighbors increasing its local clustering coefficient. Thus, by observing the distribution of $C_i(k)$ the organization of clustering with different classes of degrees can be noticed.

Table 4.4 indicates the global clustering coefficient values for the three bus transport network topologies in regular and supernode representations. Fig. 4.3 shows the histogram for the local clustering coefficient of nodes for the three bus transport networks in regular and supernode representations. It is observed from Fig. 4.3, and Table 4.4 that, due to combining of nodes in supernode representation, the local and global cohesiveness is significantly improved in all the three

Tab. 4.4: Global clustering coefficient of the three bus transport network topologies in regular and supernode representations.

	Regular	**Supernode**
Hong Kong	0.07	0.21
London	0.005	0.075
Bengaluru	0.06	0.12



(a) **(b)** **(c)**

Fig. 4.3: Histogram of local clustering coefficient for (a) Hong Kong; (b) London; and (c) Bengaluru networks in regular and supernode representations.

networks. Knowing the information about local and global cohesiveness is significant from a network analysis point of view since the local cohesiveness provides information on the available alternate path from the node during an emergency failure in its functionality, and thus acts as a great source of information in developing various routing algorithms.

4.4 Travel Distance in Hops

In a BTN, the number of hops to be traversed to accomplish a journey between any two chosen stops in a network is measured by the network parameter termed path length. Shortest path is the shortest number of links (hops) to travel between the two chosen nodes, and the average path length (geodesic path) is the average of shortest path length between all node pairs in the network given by (2.12).

Tab. 4.5: Average path length of the three bus transport network topologies in regular and supernode representations.

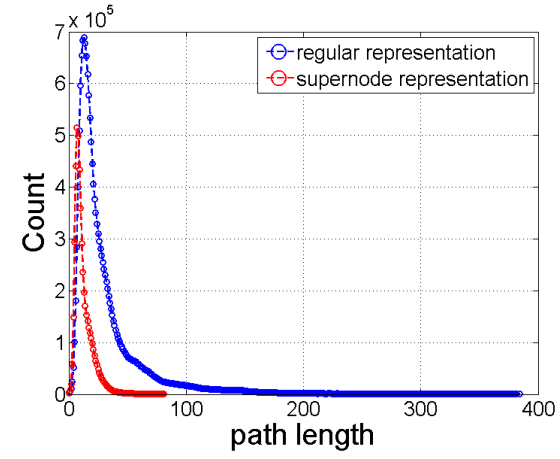
	Regular	**Supernode**
Hong Kong	**14**	**8**
London	175	102
Bengaluru	26	22

In our work, the Dijkastra's algorithm is adopted to calculate the path length between two nodes by considering the reciprocal of the edge weight to obtain the reasonable path search results [47].

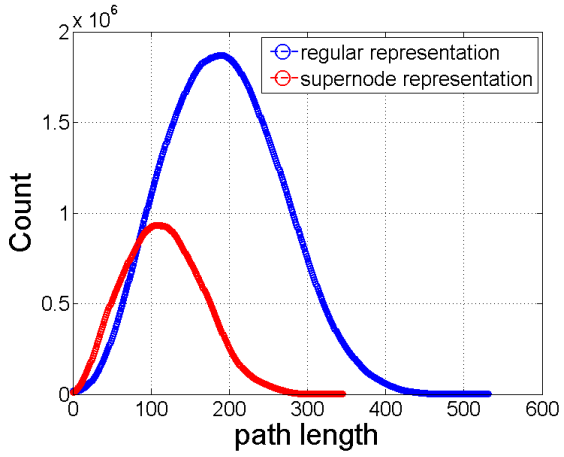
Merely representing the physical connectivity in a BTN does not typically justify the true measure of path length since there exists no physical connectivity between a few nodes in the BTN, and if such nodes can be virtually connected by a short walking distance (supernodes), the virtual connectivity will aid in effective calculation of path length. Thus, the average path length of a network is significantly affected in L-space by the existence of supernodes.

Empirically, the improvement in the perceived path length for a network with and without supernodes is illustrated in Fig. 4.4. The path length distribution is observed to follow an asymmetric unimodal distribution for Hong Kong, and a normal distribution for London and Bengaluru networks. Such distribution patterns are the virtue of inhomogeneous distribution of nodes in a city where a fewer number of stops are deployed in the suburbs/peripheries leading to long travel distances, as compared to plethora of stops at Central Business Districts (CBDs) leading to short travel distances. Table 4.5 shows the average path length values for the three bus transport networks in regular and supernode representations. However, the values in the table provides no information on the number of transfers to be undertaken as part of the journey.

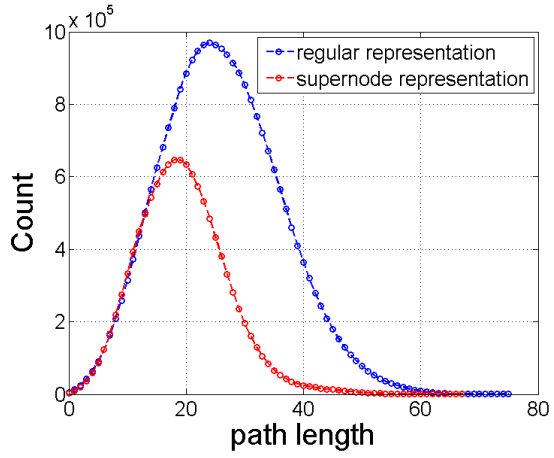
Though the supernode representation gives a more practical understanding of the average path length, without considering the total number of transfers to be made, the evaluated values seems less practical from an end user perspective. Yet, the average path length is one of the important measures in a BTN analysis



(a)



(b)



(c)

Fig. 4.4: Path length distribution for (a) Hong Kong; (b) London; and (c) Bengaluru bus transport networks in regular and supernode representations.

Tab. 4.6: Value of ω for the three BTN structures in regular and supernode representations at $p = 10^{-4}$.

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

since the number of hops is definitely a prime factor considered by the passengers in deciding a route to be taken to accomplish their journey. Lastly, average path length is also an important parameter which aids in demonstrating the small-world behavior as discussed in the next section.

4.5 Small-worldness in Bus Transport Networks

The average path length (L), and the global clustering coefficient (C) are the parameters used to quantify the structural properties of a small-world network, and is empirically evaluated using (2.15). In the literature, a lattice structure is typically considered as the initial point where $L(0)$ and $C(0)$ are evaluated, followed by which, the network is rewired with a given rewiring probability (p) to obtain $L(p)$ and $C(p)$, the small-worldness of the network is then evaluated using $L(0), C(0), L(p), and C(p)$.

In our work, the original bus network structure is considered as the initial point instead of the lattice structure since we believe that the topology of a transport network is seldom expected to be a lattice like structure, and the comparison of $C(p)$ with C_{latt} might yield in obscure results. For a fair comparison, $C(p)$ is compared with the existing network topology C_{existing}, and (2.15) is modified to (4.2).

$$\omega(p) = \frac{L_{\text{rand}}}{L(p)} - \frac{C(p)}{C_{\text{existing}}} \quad ; 0 \leq p \leq 1 \tag{4.2}$$

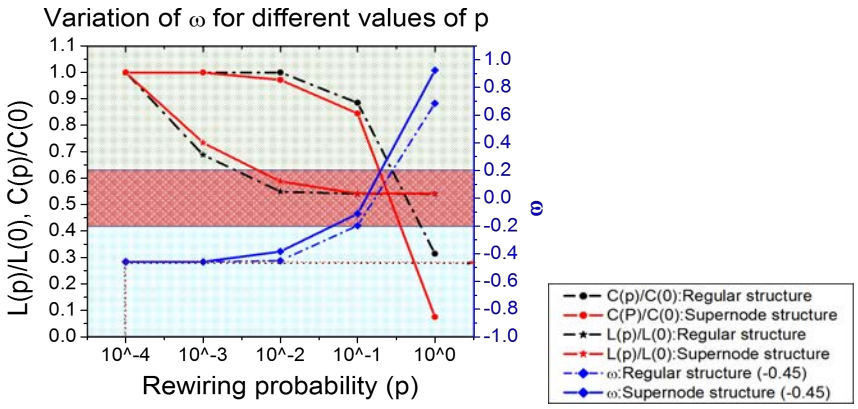
Since $C(p)$ at p=0 is same as the clustering coefficient of existing network topology C_{existing}, the term $\frac{C_{(p)}}{C_{\text{existing}}}$=1. Using the modified equation (4.2), the values of $L(0)$ and $C(0)$ are evaluated for the existing network structure (no rewiring). Then, the edges in the network are rewired with different values of rewiring probability p using the rewiring mechanism discussed in ref. [107], and

the corresponding values of $L(p)$ and $C(p)$ are evaluated for the three networks. Using the values of $L(p)$, $C(p)$, $L(0)$ and $C(0)$, the parameter ω is evaluated using (4.2), and the corresponding plots are shown in Fig. 4.5. The closer the value of ω to zero, the network behaves closer to a small-world network [52]. Table 4.6 shows the values of ω for $p = 10^{-4}$ (i.e., $p \approx 0$) for all the three BTN structures in regular and supernode representations.

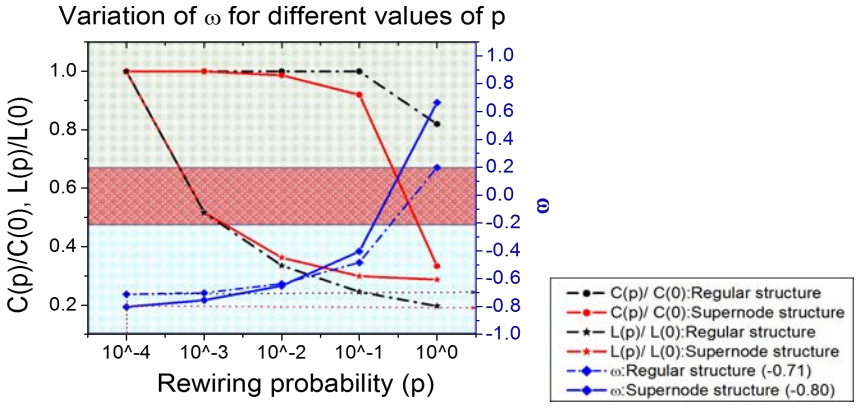
Furthermore, the plots in Fig. 4.5 indicate the variation of global clustering coefficient and average path length of a network for different values of p in regular and supernode representations. It is interesting to observe that, as p tends to one, the change in trends of L and C is better in supernode representation which leads to slower decay of curves as compared to conventional representation where the curves decay faster. The values of ω for the three networks from Table 4.6 show that Hong Kong and Bengaluru networks have the potential to behave as small-world networks with certain modifications to the existing routes. However, the value of ω for London indicates that the network requires significant modification in the routes to behave as a small-world network. However, a significant rewiring would completely change the existing network, which could be undesirable. Lastly, 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), i.e., small-worldness captures the nature of local connectivity among nodes through clustering, and captures the nature of global connectivity by measuring the minimum number of hops required to travel between two end points in a network.

4.6 Centrality Measures

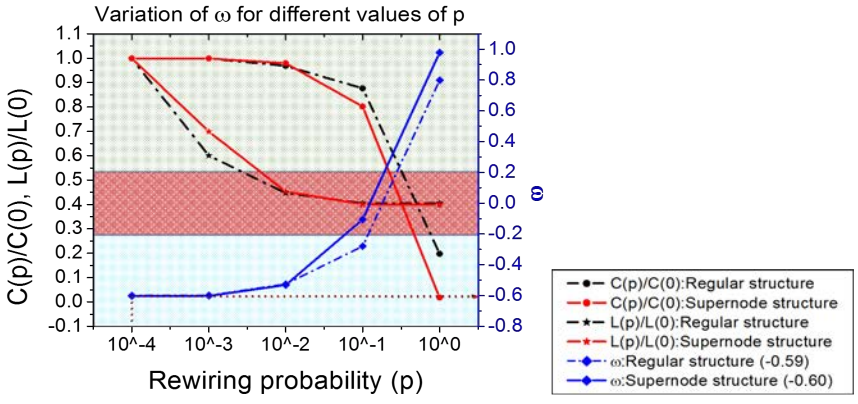
In this section, we study a few centrality measures which conveys information on a node's significance to the network based on various measures. We limit the discussion to only three centrality measures, i.e., eigen vector centrality, betweenness centrality, and hub and authority centrality, since they are found to be more relevant to the BTN analysis.



(a)

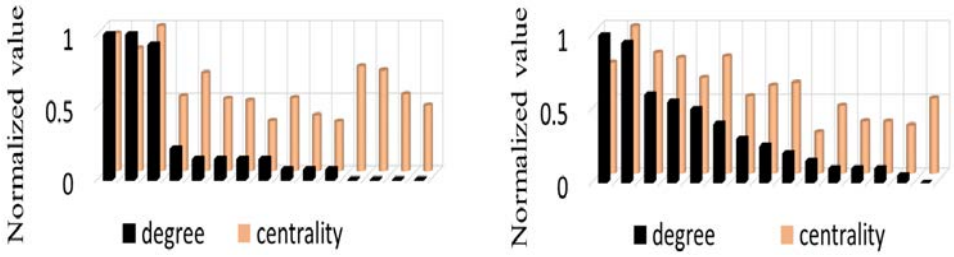


(b)



(c)

Fig. 4.5: Small-world network behavior for (a) Hong Kong; (b) London; and (c) Bengaluru networks in regular and supernode representations with $\omega(p = 10^{-4})$ highlighted.



(a) **(b)**

Fig. 4.6: Comparison of in-degree and eigen centrality values for a few nodes in Hong Kong in (a) regular; and (b) supernode representations.

4.6.1 How Influential are A Node's Neighbors?

A node is more important to the network not only because it has many neighbors (in-degree), but because it has important nodes as its neighbors, such a property is verified by eigen centrality, and is given by (2.20). Fig. 4.6 compares the in-degree and the respective eigen vector centrality of a few nodes in the Hong Kong network. It is observed from the figure that a node with high in-degree can have a low eigen centrality, indicating that the node can be less central. On the other hand, a node with low in-degree can have high eigen centrality indicating that the node can be more central, not only its in-degree is high, but also it has high degree nodes as its neighbors. That is, not all the nodes with high in-degree are central nodes in the network, or degree centrality alone might not be able to fully capture the true measure of a node's centrality. Pertaining to the real-world usage, the nodes with high in-degree and high eigen centrality caters for the maximum demand since the nodes attracts a significant influx of routes by themselves, and from their neighbors.

Furthermore, with supernode structure representation, along with procuring more practical perspective of a node's in-degree, the central nodes are more accurately predicted as a result of combining significant nodes in the network. That is, supernode representation not only improves the node connectivity but also helps to identify nodes which are having connection to more influential nodes in the network. Thus, supernodes play vital role in evaluating the true measure of a node's centrality in BTN analysis. However, the problem with eigen centrality is that, if a node has only outgoing edges and no incoming edges, the centrality of

such a node will be zero [6]. Since we encounter very few of such nodes in our analysis, eigen centrality is still conveys useful information on a node's significance to the overall network based on its connectivity to influential neighbors.

4.6.2 Hubs and Authorities in Bus Transport Networks

A hub is a node that connects to many other nodes in the network, and an authority is a node which is connected by many hubs in the network. In a BTN topology, hubs can be the nodes with high degree since they are connected to many nodes in the network, and authorities can be the nodes connected by these hubs, indicating that the authority nodes cater to the highest demand in a city. The HITS (Hyperlink-Induced Topic Search) algorithm employed to find the hubs and authorities in a network assigns every node a hub and an authority score given by:

$$h_i \propto \sum_j (u_j a_{ij}) \quad and \quad u_i \propto \sum_j (h_j a_{ji}) \tag{4.3}$$

where h_i and u_i are hub and authority scores, respectively, and a_{ij} is the adjacency matrix element corresponding to node i in the network. An unique observation is that, the hub and authority scores for a numerous nodes in the network are found to be approximately same indicating that a hub can behave as an authority, and vice versa. That is, considering the real-world usage of the network, it can be stated that the authorities are the nodes catering to highest demands, and based on our results, it can be concluded that the authority nodes in a BTN are none other than the hubs (nodes having maximum connectivity). Fig. 4.7 shows the nodes with high hub score (normalized value ≥ 0.8) in the London network in regular and supernode representations. A similar scenario is also observed for the authority scores. It is interesting to observe the effectiveness in identifying the hubs in London central business district by the HITS algorithm when the network is represented in supernode structure which the conventional graph representation failed to demonstrate.

4.6.3 Bridges in Bus Transport Networks

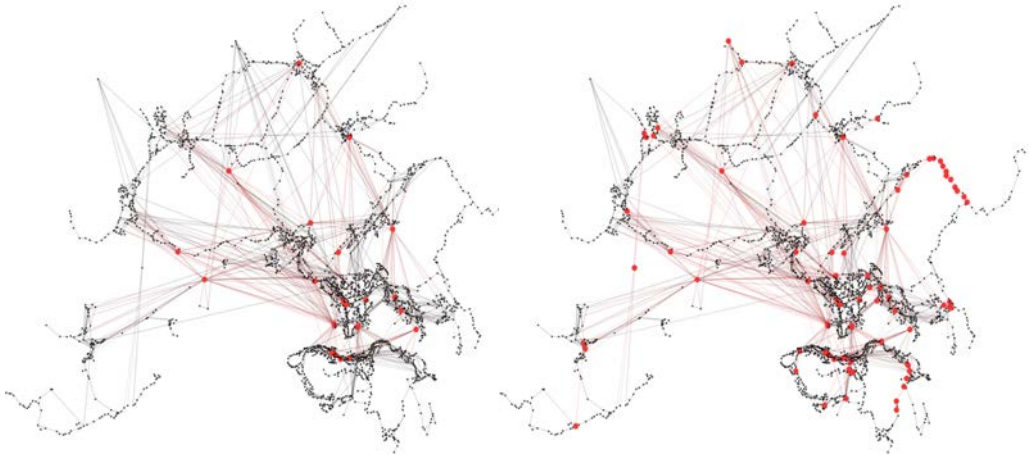
The parameter betweenness centrality highlights the significance of a node in a network considering its capability in bridging multiple shortest paths in a network,

(a) **(b)**

Fig. 4.7: The London bus network showing the nodes with high hub score in (a)regular; and (b) supernode representations.

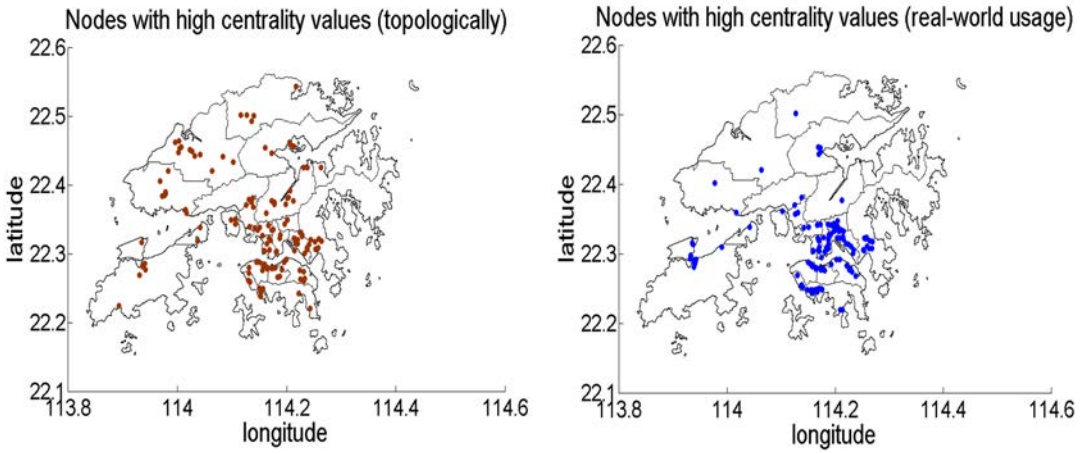
and is mathematically given by (2.16). For a given network, it is quite straight forward for a network analyst to think that the nodes with the highest degree have higher probability to serve as bridges in the network which might not be always true. For example, Fig 4.8 highlights the nodes with high degree and betweenness centrality (normalized value ≥ 0.8) in Hong Kong. From the figure it is evident that, all the high degree nodes can act as bridges in the network, whereas the vice versa is not always true.

Furthermore, combining a few significant nodes to form supernodes effectively aid in finding bridges since the supernodes compensate for the walkable catchment between the geographically closer nodes aiding in passenger interchange between the routes. Thus, though the conventional graph representation aids in finding bridges merely based on its capability in bridging multiple routes, the supernode representation further improves the chances of identifying bridges in the network since it includes both a node's capability in bridging multiple routes and a short walking distance between two nodes which supports interchange between the routes. The physical significance of this centrality is that removing these nodes from the network significantly effect the path length since these nodes potentially control the routing behaviors of both passengers and buses in the network.



(a) **(b)**

Fig. 4.8: Bus stops in Bengaluru network with high betweenness centrality (normalized value ≥ 0.8) in (a) regular; and (b) supernode representations.



(a) **(b)**

Fig. 4.9: Hong Kong bus transport network in supernode representation with highly central nodes (a) evaluated using different centrality measures; and (b) evaluated using static demand estimation method.

4.7 Figure of Merit

In this section, we compare the topological centrality of a few nodes with their geographical centrality. That is, a set of 130 nodes are identified in the Hong Kong BTN structure with high values (normalized value > 0.8) of degree-, hub-, betweenness- and eigen- centralities. These 10 nodes are considered highly central nodes in the network pertaining to topological analysis. On the other hand, for the same set of 130 nodes the node weight value evaluated using (3.8) are extracted. By comparing a node's topological centrality with its geographical centrality we observe that among the chosen 130 nodes almost 80 nodes are found to be both topologically significant and geographically important which resulted in a 63% figure of merit. Fig. 4.9 shows the comparison of topologically central nodes versus geographically significant nodes for the chosen 130 nodes in Hong Kong BTN.

Based on our observation, we can state that the geographical significance of the node evaluated at zonal level using (3.8) yields 60% same result as estimating different topological centrality measures at the global level. Thus, the strong correlation between the topologically central nodes and the geographically central nodes reveals that the proposed static demand estimation method for assigning node weight aids in better identification of influential nodes in the network. Whereas, the 40% discrepancy among the chosen 130 nodes act as great source of information to the operators to further modify and optimize the routes to offer improved and more sustained services. The figure of merit is currently evaluated for only Hong Kong bus transport network. The similar approach can be employed for London and Bengaluru networks, which can be accomplished as part of future work.

4.8 How Efficient is A Bus Transport Network Topology?

In this section, we analyze the topological efficiency of a network in terms of the distance traveled in hops, which is a conventional way to quantify the topological efficiency. We then remodel the definition of topological efficiency in terms of time rather than the distance which is useful from a passenger perspective. As

Tab. 4.7: Comparison of topological efficiency of the three BTN structures in regular and supernode representations.

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

per network theory, topological efficiency of a network is given by

$$\eta_G = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{4.4}$$

where d_{ij} is the shortest path between nodes i and j, and N is the network size. Table 4.7 shows the comparison of topological efficiency for the three networks in regular and supernode representations. From the Table 4.7, it is evident that the Hong Kong network is topologically more efficient than the London and Bengaluru networks in both regular and supernode representations. Also, the efficiency of the three networks is higher in supernode representation as compared to the regular representation. In fact, we can state that the true measure of topological efficiency is better exploited in supernode representation than the conventional representation. Although no physical change is made to the actual network, a slight restructuring in network representation aids in gathering insightful information on the actual behavior.

Equation (4.4) conveys information on whether the network offers shorter distance of travel between any two given nodes. However, from a passenger standpoint, end-to-end travel delay seems a more reasonable metric than the distance. Of course, other parameters like minimum transfer, cost effectiveness and convenience are also to be considered. In our work, (4.4) is remodeled in terms of delay, and is given by

$$\eta_{G,t} = \frac{1}{N(N-1)} \sum_{\substack{i=1\ldots n-1 \\ j=i+1\ldots n}} \frac{d_{ij}}{v_{ij}} \tag{4.5}$$

where d_{ij} measures the actual geographic distance of every hop along the shortest distance between nodes i and j. v_{ij} is the maximum speed attained along every hop of the shortest path between the nodes i and j. The term $\frac{d_{ij}}{v_{ij}}$ gives the

end-to-end travel delay. However, obtaining the data about v_{ij} from real-world experiments is the main obstacle in evaluating $\eta_{G,t}$. Thus, to validate equation (4.5), a simple simulation is accomplished using SUMO (Simulation of Urban Mobility) tool for a specific route in Hong Kong, which is supported by the real-world data from KMB (Kowloon Motor Bus Co.) operator in Hong Kong. The same approach can be scaled up for all the routes to obtain $\eta_{G,t}$ for the overall network.

4.8.1 SUMO Simulations

SUMO is a microscopic multi-modal traffic simulator which allows the behavior of each vehicle to be explicitly controlled and monitored [108]. In this section, the details of the simple simulation conducted using SUMO tool are discussed.

Step 1. Building a road network topology:

To build a road network in SUMO, we can either import the network from OpenStreetMap [109], or build the network manually. In our simulation, the network topology is built manually where The bus stops are treated as nodes, and the road segment connecting the two bus stops is treated as an edge. The information about the location of bus stop, # lanes per road segment, the pedestrian crossing, road junction, and traffic light are extracted from the OpenStreetMap. All the traffic lights are generated with a default cycle of 90 s (40 s green, 5 s amber, 40 s red followed by 5 s amber to switch to the next cycle of green). The simulation considers the road topology as described above for a specific bus route (Route #1) operating between the source node S (CHUK YUEN ESTATE), and the destination node D (STAR FERRY). The end nodes S and D, and the specific route between them is chosen for the simulation since the route passes through different zones in the city with significant POIs. Such information will help us to verify the node weight analysis as discussed in Section 3.3.1 in Chapter 3. The real-world dataset from the KMB operator in Hong Kong is also obtained for the same route #1.

Step 2. Setting up the routes for buses and other general vehicles in the network:

The route details and the frequencies for the bus routes are configured according to the official timetable of KMB [110]. We use "activitygen" to generate the

traffic other than the buses [111]. Activitygen is a tool in the SUMO simulator that generates traffic in a network based on the activity in a zone, and is termed activity based demand generator. An activity can be regarded as a trip going to or from an office, school, or free time travel. By providing the input data on the number and locations of POIs, the number of people living in the zone, the number of people traveling from other zones to the chosen zone, the working hours in a day, etc., activitygen generates activities happening in the zone. Using another tool called Duarouter [112], every activity in the zone is assigned a route based on the shortest path between the source and the destination of the activity. The destination could be a POI that the passenger wants to reach, and the source could be a location within the zone or outside the zone where a passenger starts the journey. If a passenger starts the journey beyond the chosen zone, a specific location through which he enters the given zone can be explicitly configured in SUMO. The routes generated by Duarouter are for the generic vehicles (other than buses). Fig. 4.10a shows a snapshot of the simulation setup described above, indicating the bus stops, traffic lights, pedestrian walkways and POIs.

Step 3. Run the simulation:

For simulation purpose, the road network as described in step 1 is employed, and the route set up as described in step 2 is employed, and the simulation is run for a duration of 3600 s for the morning peak hour (8:00 to 9:00 am).

Step 4. Extract the results from SUMO:

Since SUMO is a microscopic simulator, it logs the results of individual vehicles in a trace file with a sampling period of 1 s. Additional information like the geographical position of a vehicle, the route information, lane information, waiting period at signals, and the velocity of a vehicle can also be found in the trace files. Using this trace file, the time mean speed of the vehicles at every 1 sec is extracted for every road segment between S and D, from which the maximum speed achieved along a road segment is evaluated.

Step 5. Calculate the end-to-end travel delay:

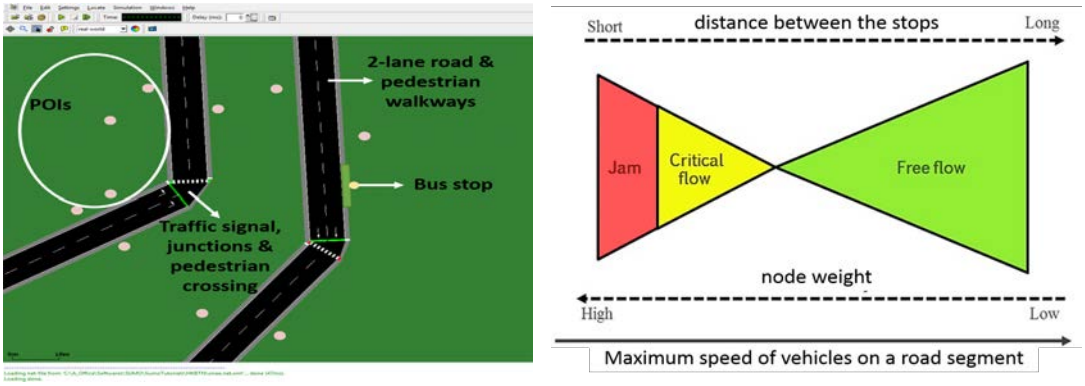
Tab. 4.8: Comparison of end-to-end travel delay between chosen nodes S and D during morning peak hour.

	HKeTransport	**Google Maps**	**Simulation**	**Empirical**
end-to-end delay (min)	65	57	56.5	52

The end-to-end travel delay is calculated between the two chosen nodes S and D using (4.5), where the geographic distance between the bus stops is calculated using the Haversine formula [102], and the speed is calculated as described in step 4. The final result of our simulation is compared with the results from Google map's, the Hong Kong's e-Transport application, and the empirical result from the real-world KMB dataset, and is tabulated in Table 4.8. It is observed from the table that the results from various sources have a small delta difference which seems reasonable in a real-world. Thus, a simple and semi-realistic simulation in SUMO aided in evaluating the topological efficiency for a particular route, which can be scaled-up for the other routes in the network.

An important observation that followed the simulation is the dependency of the vehicular speed along a road segment on the node weight. From the simulation results, we observe that, as the number of POIs around a particular bus stop increases, the maximum speed attained by the vehicles along the particular road segment decreases for the peak hour simulation. The situation goes worst when the distance between the bus stops decreases, and the node weight increases, i.e., as shown in Fig. 4.10b, the speed of the traffic changes from free flow to critical flow, and finally to a jammed state as the node weight increases, and the distance between the bus stops decreases.

To support the simulation results, a real-world dataset provided from the KMB operator in Hong Kong is employed. The data was collected for a period of one week during October 2017 for the morning (8:00-11:00 AM) and evening peak hours (4:00-7:00PM). The data x (latitude), y (longitude), t (time), v (velocity at a given time) is sampled with a sampling rate of one second along the chosen route (Route #1). Using the data, the maximum speed achieved along a road



(a) **(b)**

Fig. 4.10: (a) Snapshot of the SUMO simulator; and (b) generalized flow prediction for dependency of vehicular speed on POI density and distance between the stops.

segment is extracted for both the morning and evening peak hours considering an average of 30 trips along the chosen route (a road segment is considered as a segment between successive bus stops, i.e., it can be a collection of smaller road segments separated by road junctions or traffic lights).

Fig. 4.11 shows the dependency of the maximum speed attained along a road segment ($V_{\max}$) on the normalized node weight (w_{i_norm}). A comparison of the simulation and empirical data showed that, as the node weight increases, the $V_{\max}$ decreases. This condition might not be true always, unless the distance between the successive nodes is considered. Fig. 4.12 plots the distance between two successive stops (d_{ij}), the normalized node weight (w_{i_norm}) and the maximum speed achieved along a road segment ($V_{\max}$) for the morning and evening peak hours for route #1. From Fig. 4.12, we see that bus stops no. 9, 10 and 11 (S9, S10, S11) have relatively higher node weight. However, since stops S9 and S10 are geographically closer, the $V_{\max}$ along the road segment is smaller as compared to the road segment between the stops S10 and S11. A similar scenario is observed between stops S17, S18 and S19. Thus, from our findings we can infer that, with increasing node weight (demand) and geographically closer bus stops, the maximum speed achieved along the road segment reduces significantly.

It is acceptable that the operators deploy more stops to meet the greater demand, however, the operators should note that closer bus stops would eventually lead to a state of traffic congestion. Hence, the route planning and the stop

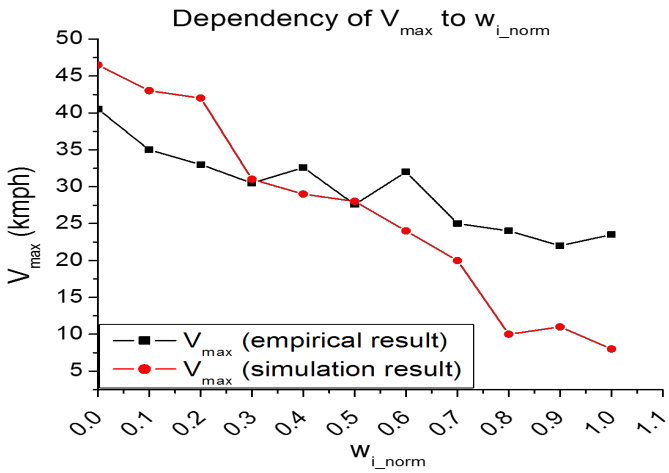


Fig. 4.11: Comparison of the simulation and empirical results for the dependency of vehicular speed on the node weight.

deployment needs judicious design to overcome the congested travels. Though the simple simulation carried out in Section 4.8.1 is primarily to get an idea on the analysis of topological efficiency in terms of the end-to-end travel delay, the simulation provided additional information on the dependency of $V_{\max}$ on w_{i_norm} and d_{ij}.

4.9 Summary

Finally, the analysis of the topological properties and structural behavior of the three bus transport networks in both regular and supernode representations revealed interesting information pertaining to node's connectivity, network cohesiveness, minimum travel distance in hops, nodes having influential neighbors, and the capacity of a node in bridging multiple routes. At every stage of the BTN analysis, it was observed that supernode representation offered a more practical understanding of the inherent network behaviors as compared to the conventional graph representation. Though no significant changes were made to the actual topology, a slight restructuring offered insightful results on the network behaviors which were found to be useful from both the operator and passenger perspectives.

Public transport networks (PTN) contain multiple layers of traffic carrying networks such as bus-, subway-, metro-, tram-, ferry- networks, etc. Existing

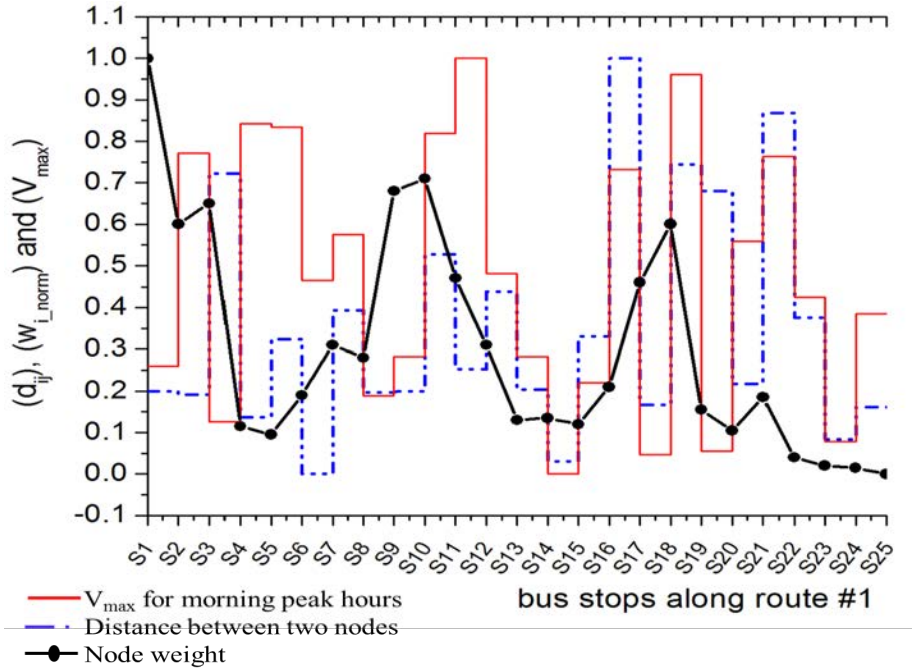


Fig. 4.12: Dependency of the maximum speed achieved along a road segment ($V_{\max}$) on the distance between two successive stops (d_{ij}) and the node weight (w_{i_norm}) for the morning and evening peak hours along the bus route #1. All the values are normalized to ensure data integrity.

works on PTN analysis using graph theory have emphasized on either the individual transport networks by considering them as isolated mono-layer networks, or the non-interacting aggregated multi-layer networks. Though the different transport layers share common features when analyzed as individual mono-layer structures, by understanding the interconnectedness among different mono-layers, a more meaningful insight is gained into the overall network structure and its dynamics. Also, since passengers use multiple transport modes (on different transport layers) to reach their destinations, it is of practical importance, though rarely considered, to study the interaction and connectivity between network layers of different transport modes as discussed in the next chapter.

Chapter 5

Multi-layer Analysis

This chapter focuses on demonstrating the need for considering the interdependency of multiple transport layers to capture the true contribution of a node and edge to the overall network. The term multi-layer in the current chapter indicates the consideration of two transport modes, i.e., the bus and metro transports.

The topological representation of bus transport network follows a directed and weighted graph in supernode representation, whereas, for the metro transport network, an undirected and weighted graph in conventional graph representation is considered since the existence of supernodes are less probable in metro networks. The major contributions of the chapter are two fold:

1. A *spatial amalgamation* method is proposed to integrate the two non-interacting mono-layers (BTN and MTN), and represent them as an interacting and integrated multi-layer.

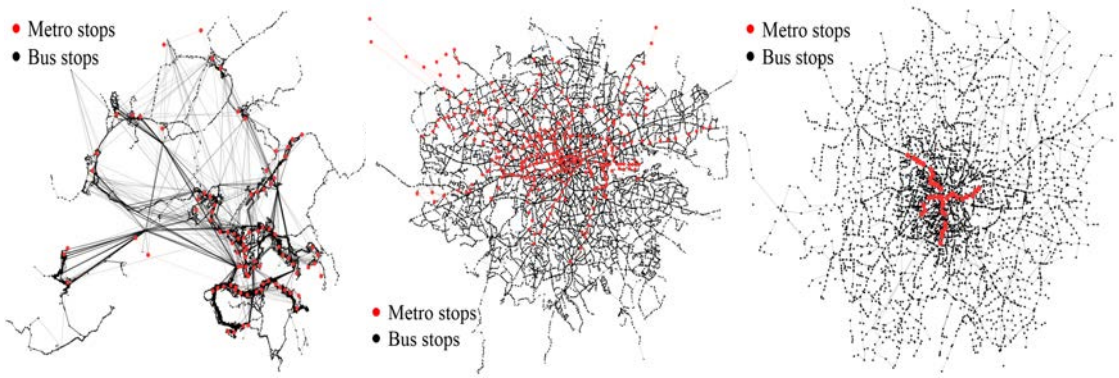
2. An extension of the node weight approach for multi-layer is proposed where, a node wight is assigned to a node by ignoring and considering the interdependency of the multiple transport modes.

5.1 Multi-layer Network

A multi-layer network M is defined by [67],

$$M = \left(V_a, E_a, \tilde{V}, \mathrm{l}\right) \tag{5.1}$$

where $\tilde{V}$ is the node set containing both regular nodes and supernodes inclusive of all the layers. $\mathrm{l} = \{l_a\}_{a=1}^{d}$ is the set of elementary layers defined by d aspects



(a) **(b)** **(c)**

Fig. 5.1: Independent BTN and MTN mono-layer topologies for (a) Hong Kong; (b) London; and (c) Bengaluru cities.

or dimensions such that there is one elementary layer set l_a for each aspect d. For $d = 1$ (single aspect), the multilayer network reduces to a mono-layer network. In our work, $d = 2$, with an elementary layer and an additional layer. Also, $V_a \subseteq \tilde{V}$ is a node set associated with different layers l_1 to l_d such that $V_a \times l_1, \times l_2, .., \times l_d$. $E_a \subseteq V_a \times V_a$ is the edge set containing both regular edges and superedges inclusive of all the layers. Among the BTN and MTN layers, we chose BTN layer as the elementary layer and MTN layer as the additional layer (which can be interchanged). Fig 5.1 shows the topology of the non-interacting mono-layers (BTN and MTN) for Hong Kong, London, and Bengaluru cities. In the future sections, the BTN layer is treated as layer-α, and MTN layer is treated as layer-β.

An important feature in transport networks is that they belong to the category of layer-disjoint networks where a node or edge exists in at most one layer [66].

$$(n_i)_\alpha \in V_\alpha, \;\; (n_i)_\beta \in V_\beta \tag{5.2}$$

where a node n_i is present either in layer α (BTN layer) or β (MTN layer). The layer-disjoint property signifies an important observation that there exists no edges between the two layers in the actual network structure, and the two layers are typically connected by a small walking distance. Hence, to integrate the two layers, in our work, the spatial amalgamation method is employed which makes use of the walkable catchment between the multiple transport modes, the details of which are discussed in next section.

5.1.1 Spatial Amalgamation Approach

In spatial amalgamation approach, the two non-interacting mono-layer networks are integrated and represented as interacting multi-layer network using the idea of walkable catchment. A geographical area with a radius of 500 m is considered as the walkable catchment considering it to be a maximum distance the passengers would prefer to walk for interchanging between the two transport modes. That is, in spatial amalgamation method, a geographical area with a radius of 500 m around a node in MTN layer is marked, the nodes in the BTN layer that falls within the region of walkable catchment are extracted, and are termed as the intermediate layer nodes or coupling layer nodes. A feature in ArcGIS tool called *buffer* is employed to accomplish the task. Considering the actual physical infrastructure, the nodes in the intermediate layer are the stations that act as transfer points for the passengers opting multiple transport modes to accomplish their journey.

As an example, Fig. 5.2a shows the layer α, layer β, and the coupling layer for the London network. Fig. 5.2b demonstrates the spatial amalgamation approach employed to integrate the two non-interacting mono-layers. The set of nodes in the intermediate layer is a subset of nodes in BTN layer that aids in establishing the virtual connectivity between the BTN and MTN layers. According to the explanation above, the intra-layer edge set E_α and E_β, and the inter-layer (coupling layer) edge set E_C are mathematically defined as

$$E_\alpha = \{e_{ij}\} \mid e_{ij} \rightarrow (n_i, n_j)_\alpha \ \forall n_i, n_j \in V_\alpha \tag{5.3}$$

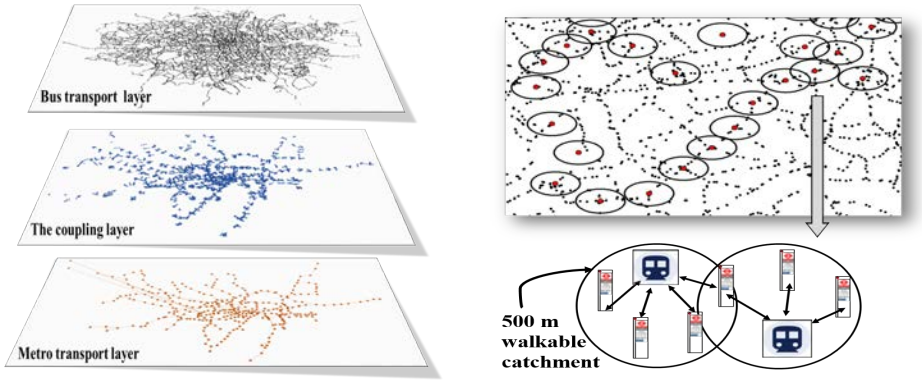
$$E_\beta = \{e_{kl}\} \mid e_{kl} \rightarrow (n_k, n_l)_\beta \ \forall n_k, n_l \in V_\beta \tag{5.4}$$

$$E_C = \{\tilde{e}_{ik}\} \mid \tilde{e}_{ik} \leftrightarrow (n_i, n_k), n_i \in V_\mathrm{C}, n_k \in V_\beta \tag{5.5}$$

where α and β are the BTN and MTN layers, respectively; V_β is the node set of layer β; and V_C is the set of overlapped nodes in the 500 m walkable catchment, such that $V_C \subseteq V_\alpha$.

5.1.2 Node Weight Approach for Multi-Layer Network

In this section, we use the node weight approach discussed in Section 3.3.1 in Chapter 3 to assign a weight to a node, with and without considering the interaction between the multiple layers. Using the node weight approach, a set



(a) **(b)**

Fig. 5.2: (a) The layers α(BTN), β(MTN) and coupling layer in the London multi-layer network; (b) an example demonstrating the spatial amalgamation approach employed to integrate the two non-interacting mono-layers.

of influential nodes is determined that quantify their significance to the overall network, and also to the mono-layer to which the nodes are associated with.

To suit the node weight approach for multi-layer network, the equations discussed in Section 3.3.1 are modified as below:

$$(\mathrm{NOP}_{i,\alpha})_z = (\rho_{P_\alpha}/\rho_{N_\alpha})_z \tag{5.6}$$

$$(\mathrm{NOP}_{\mathrm{j},\beta})_z = \left(\rho_{P_\beta}/\rho_{N_\beta}\right)_z \tag{5.7}$$

where, $\mathrm{NOP}_{\mathrm{i},\alpha}$ is the node occupying probability of a station i in layer α, ρ_{P_α} is the static population density accessing the stations in layer α in a zone z, and ρ_{N_α} is the node density in layer α in a zone z. For simplicity, the term *zone* in Section 3.3.1 is represented as z, $\forall z = 1, 2, .., Z$. The notations follow similarly for layer β.

The information regarding the static population accessing the nodes in a specific layer is derived from the statistical data of the total population count in a zone [113, 114, 115].

$$(P_\alpha = c_4 * P_T)_z \tag{5.8}$$

$$(P_\beta = c_5 * P_T)_z \tag{5.9}$$

where P_α and P_β denote the population count accessing the stations in layer α and β respectively, P_T denotes the total static population in a given zone. Table 5.1 indicates the value of c_4 and c_5 for the three cities. Using P_α and P_β, we determine ρ_{P_α} and ρ_{P_β}.

Tab. 5.1: Population count accessing a specific layer is derived from the total population as given in 5.8 and 5.9.

	Hong Kong	**London**	**Bengaluru**
BTN (c_4)	23%	21%	58%
MTN (c_5)	77%	79%	42%

Thus, the node weight approach for non-interacting mono-layer is given by:

$$(w_{i_\alpha})_z = (\frac{\rho_m + \rho_{P_\alpha}}{\rho_{N_\alpha}})_z + k_{i_\alpha} \tag{5.10}$$

$$\left(w_{j_\beta}\right)_z = (\frac{\rho_m + \rho_{P_\beta}}{\rho_{N_\beta}})_z + k_{j_\beta} \tag{5.11}$$

where $(w_{i_\alpha})_z$ is the weight of a node i in layer α in a chosen zone z, and k_{i_α} is the node degree which indicates the connectivity of a node in layer α. The notations follow similarly for layer β. The node weights are normalized to ensure the data integrity in all the zones. The closer the value of normalized node weight to one, more influential is a node.

Considering the interdependency of multiple transport modes, the node weight approach for the multi-layer network is given by,

$$w_i = (w_{i_\alpha})_z + C_{b_i} \tag{5.12}$$

$$w_j = (w_{j_\beta})_z + C_{b_j} \tag{5.13}$$

where w_i is the overall weight of a node i, $(w_{i_\alpha})_z$ is the weight of a node without considering the interaction between the two layers as given by (5.10), and C_{b_i} is the betweenness centrality of node i as discussed in Section 4.6.3.

A node's capacity in bridging multiple routes after integrating the two layers conveys its true significance to the overall network due to which the parameter betweenness centrality is employed in assigning node weight. Figs. 5.3a, 5.3c, 5.3e indicate the set of influential nodes according to non-interacting mono-layer analysis, and Figs. 5.3b, 5.3d, 5.3f indicate the set of influential nodes according to interdependent multi-layer analysis. It is observed from the figures that a node's significance to the overall network differs significantly by considering and ignoring interdependency between the layers. This shows that neglecting the

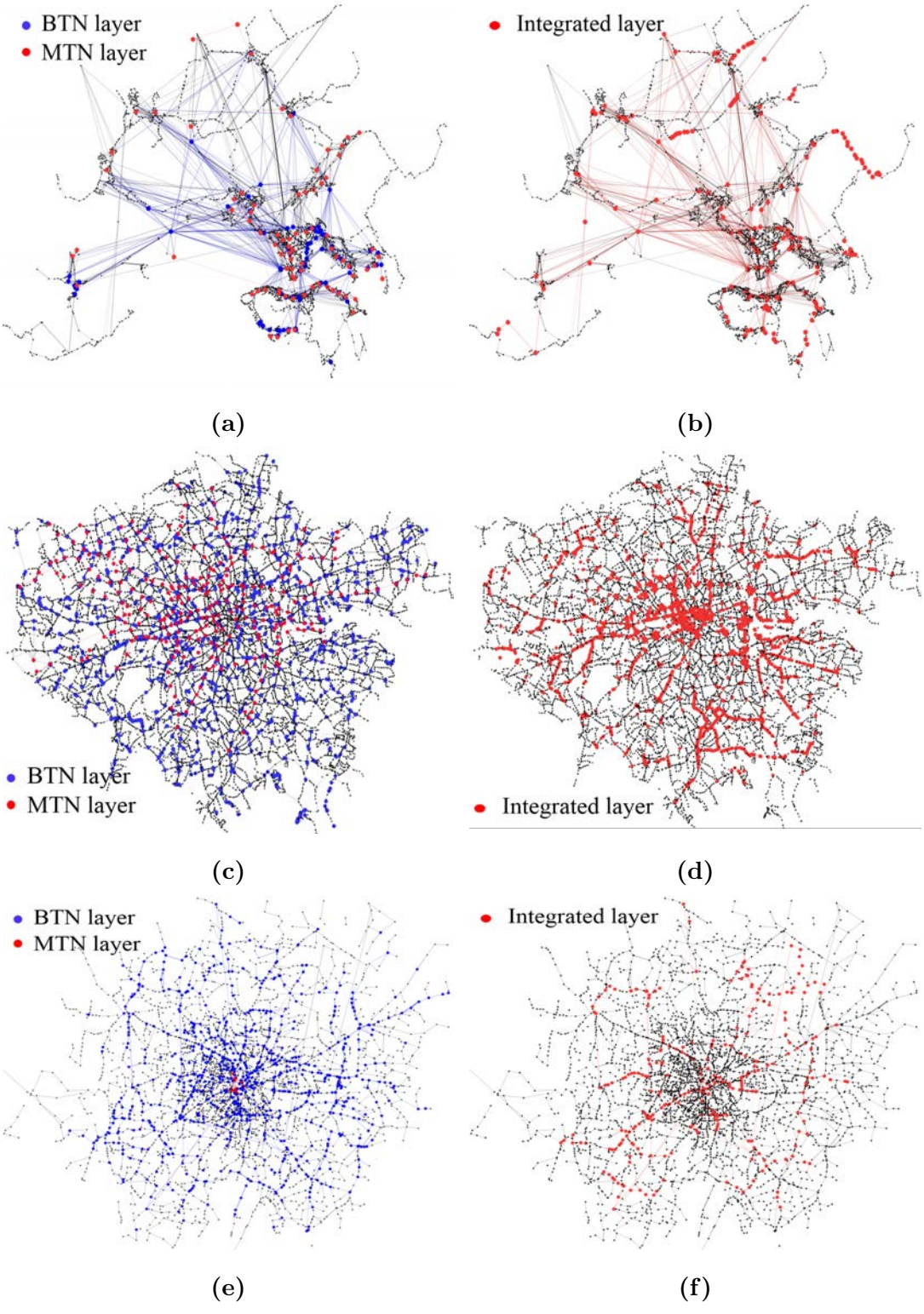


Fig. 5.3: (a) The influential nodes ($w_i \geq 0.8$) in non-interacting BTN and MTN mono-layers, and the integrated multi-layer for (a,b) Hong Kong; (c,d) London; and (e,f) Bengaluru network topologies.

interaction between the transport layers will completely bias our understanding of the overall network behavior.

5.2 Summary

To simultaneously quantify a node's significance to the overall network, we first considered a node's significance in the isolated mono-layer, then, we considered its significance in the interacting multi-layer. The ranking of nodes by considering and ignoring the layer dependency conveyed the information that ignoring the layer dependency will yield in absurd results. Though the results in the current chapter are primarily based on the simple approach of spatial amalgamation, knowledge about additional factors like variable walkable catchment and realistic passenger distribution will further improve the understanding of layer interdependency.

In the literature, there exist a numerous works whose contributions are limited to only mono-layer analysis which might cater partial results to the network operators. Therefore, to offer more meaningful suggestions to the operators, the consideration of interdependency among multiple layers is always suggested. However, the huge complexity involved in considering a numerous real-world factors to integrate the multiple layers can be a primary reason to ignore the layer interdependency in the previous works. Until this point, we emphasized on the PTN analysis based on the concepts of graph theory, where the topological behavior of multiple transport modes and their interaction are studied to cater useful information to the network operators to offer better service to the passengers.

Chapter 6

Impact of PTN Topology on Vehicular Network Connectivity

In this chapter we aim to integrate the two different disciplines of study: PTN analysis using graph theory and the VANET connectivity analysis, to understand the dependency of vehicular network connectivity on transport network topology, from a macroscopic perspective. In VANETs, the communication can be either vehicle-to-vehicle (V2V), or vehicle-to-infrastructure (V2I). In our work, we restrict our analysis to V2I which is seldom explored in the literature.

In the V2I connectivity analysis, the vehicles are buses, and the infrastructure is a bus stop which is equipped with a communication device to support information exchange. The vehicles considered are limited to only buses since the mobility pattern of the buses is significantly different and partially deterministic (due to constrained and structured mobility). Contact duration, a key parameter in VANET connectivity analysis that determines the throughput of the information exchange is employed to assess the nature of V2I connectivity. The major contributions of the chapter are four fold:

1. We simulate a semi-realistic scenario in SUMO tool to obtain the synthetic mobility traces of the buses which are employed to evaluate the V2I contact duration as discussed in Section 6.4.

2. For the microscopic analysis, we consider two topological metrics: weighted in-degree and demand centrality of a node to demonstrate the nature of V2I contact duration.

3. For the macroscopic analysis, the results obtained in Step 2 are used to

assess the distribution pattern of contact duration for the two topologies: scale-free and random network topologies.

4. Lastly, the robustness of the VANET communication network is assessed by the random node removal approach to determine the network's ability to sustain during malfunctions.

6.1 Transport Network Topology and The Vehicular Network Connectivity

In its first kind, we aim to integrate two different fields of study, namely, PTN analysis using graph theory and connectivity analysis in VANET, to study the impact of the transport network topology on the vehicular network connectivity, from a macroscopic perspective. To accomplish the study, initially, we assess the nature of V2I connectivity at a microscopic level, later, the microscopic results are used to demonstrate the nature of V2I connectivity at a macroscopic level. To accomplish the study, initially, we assess the nature of V2I connectivity at a microscopic level, later, the microscopic results are used to demonstrate the nature of V2I connectivity at a macroscopic level.

Specifically, we consider two topological metrics namely, weighted in-degree and demand centrality of a node that were discussed in Chapter 2 and 3. The two topological metrics are employed to demonstrate the nature of V2I connectivity, which we term as microscopic connectivity analysis. Later, by studying the distribution pattern of the two topological metrics for a given transport network topology, we determine the distribution pattern of V2I contact duration, which we term as macroscopic connectivity analysis. Fig. 6.1 demonstrates the details of the approach employed from both micro- and macroscopic perspectives.

The scale-free and random network topologies are practiced in the current chapter since the details on the behavior of the two topologies were previously discussed in Chapter 2 and 4. Recollecting our results from Chapter 4, the Hong Kong network demonstrated the scale-free property in supernode representation, while the London and Bengaluru networks demonstrated random topological behavior in both regular and supernode representations. In this chapter, we consider the results concerning to Hong Kong and London network topologies in supernode

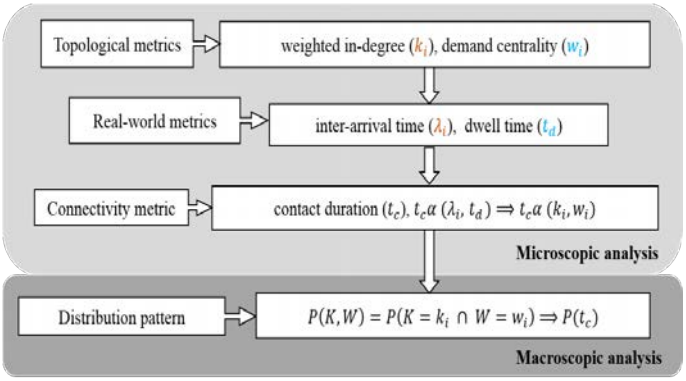


Fig. 6.1: The flow diagram indicating the consideration of the two topological parameters that determine the nature of V2I contact duration (the definition of individual parameters are discussed in Section 6.3).

representation, and demonstrate the impact of the two topologies on the nature of vehicular network connectivity.

6.2 Topological Metrics: A Brief Review

Weighted in-degree:

As discussed in Section 4.1 in Chapter 4, the connectivity pattern of a node is evaluated by a network parameter termed degree (k). The degree of a node signifies the number of edges incident on it (i.e., the number of bus stops a given bus stop is connected to). Whereas, a weighted node degree indicates the number of routes servicing a station. For a directed graph, the weighted node degree (k_{w_i}) is given by

$$k_{w_i}^{\text{in}} = \sum_j a_{ji} w_{ji}, \quad k_{w_i}^{\text{out}} = \sum_j a_{ij} w_{ij}, \quad k_{w_i}^{\text{total}} = k_{w_i}^{\text{in}} + k_{w_i}^{\text{out}} \tag{6.1}$$

where a_{ij} is the adjacency matrix element corresponding to the i^{th} node, and w_{ij} is the weight, i.e., the number of routes servicing a station.

In our work, the weighted in-degree of a node is considered as the first topological metric since it conveys information on the number of routes servicing a bus stop. That is, weighted in-degree of a node determines the number of buses arriving at a bus stop, which in turn determines the density of buses around the bus stop, and thus contributes significantly in evaluating the V2I contact duration. In the future sections, we denote $k_{w_i}^{\text{in}}$ as k_i for simplicity.

Demand centrality:

The demand centrality (node weight) of a node is considered as the second topological metric, the details of which were discussed in Section 3.3.1 in Chapter 3, and is given by

$$(w_i)_{\text{zone}} = \left(\frac{c_1 \sum_{m=1}^{4} \rho_m + c_2 \rho_p}{\rho_N} \right)_{\text{zone}} + c_3 k_i \quad ; \text{zone} = 1, 2, ..., z \tag{6.2}$$

$$(w_{i\_\text{norm}})_{\text{zone}} = \left(\frac{w_i - w_{\min}}{w_{\max} - w_{\min}} \right)_{\text{zone}} \quad ; \text{zone} = 1, 2, ..., z \tag{6.3}$$

where the demand centrality of a node conveys information on the NOP (i.e., number of people accessing a station), which determines the dwell time of a bus at the bus stop, and thus contributes significantly in evaluating the V2I contact duration. Higher the node weight, the demand serviced by the node is high, and thus, the dwell time of the buses is high. For simplicity, the parameter w_{i_norm} is represented as w_i in the future sections.

Thus, the two topological metrics: weighted in-degree and demand centrality of a node determines the density and dwell time of the buses at a bus stop, respectively, and contributes significantly in evaluating the V2I contact duration (t_c).

6.3 Definitions

In this section, we discuss the definition of a few parameters which are found to be useful in the future sections.

1. **Dwell time (t_d)**: the time spent by the buses to board or alight the passengers at a designated bus stop. Typically, the dwell time of the buses at the bus stop depends on the number of passengers accessing it (NOP), which is determined by the node weight. Thus, we determine the dwell time based on the demand serviced by a node, i.e., higher the node weight, the demand serviced by the node is high, and thus, the dwell time of the buses is high.

2. **Vehicle-to-infrastructure contact duration (t_c)**: The term infrastructure refers to a bus stop equipped with wireless communication device which

can support information exchange for a specific range R_I. The vehicle-to-infrastructure contact duration is the time period during which a single hop communication is possible between the buses and the bus stop. We set a transmission range of $R_v = 100$ m for the vehicles, and $R_I = 100$ m for the bus stops (or infrastructure) as shown in Fig. 6.2.

The main aim of evaluating the nature of vehicle-to-infrastructure (V2I) contact duration is that the buses reach a state of immobility while nearing the bus stop (to board and alight the passengers) which offers an increased and sustained connectivity between vehicles and infrastructure. The V2I contact duration is mathematically defined as

$$t_c = t_{\text{start}} - t_{\text{end}} \tag{6.4}$$

where t_{start} is the time instant at which the infrastructure detects a bus when it enters the range R_I, t_{end} is the time instant at which the infrastructure fails to detect the bus when it leaves the range R_I. The two time stamps are extracted based on the footprints generated by SUMO simulation results as will be discussed in Section 6.4. For example, in Fig. 6.2, buses 1 and 2 contributes towards evaluation of t_c since they are within the range R_I, however, buses 3 and 4 do not.

3. **Inter-arrival time or time headway (λ_i)**: is the time interval between two successive bus arrivals for the same route. For a given node with k_i = m, the inter-arrival time is a set of elements with each element denoting the time headway of individual route, i.e., $\lambda_i = \{\lambda_i^1, \lambda_i^2, \lambda_i^3, ..., \lambda_i^m\}$. Furthermore, We assume that the arrival rate of buses at a bus stop follow Poisson distribution with mean = λ_i, and standard deviation =0. That is, with Poisson arrival, we assume that the buses arrive at the station on time, however, if the berth is occupied, the buses enter waiting state until the berth is cleared. During this waiting period, if the buses are within the range R_I, they contribute towards t_c. The value of λ_i is randomly chosen such that $4 \leq \lambda_i \leq 10$ (the time headway of the buses vary between 4 to 10 min) for the peak hour simulation which is a reasonable assumption valid in real-world scenario.

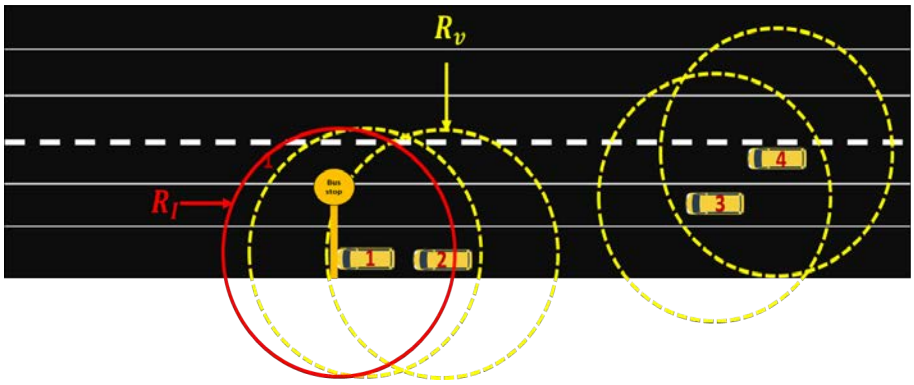


Fig. 6.2: The details of the simulation setup along with the required parameters.

Note: The background traffic is not considered in our simulation. Only one hop connectivity of buses is considered, i.e., the contact duration indicates that the buses will be able to communicate with the infrastructure directly. Only east bound vehicles are considered in the simulation. We assume that the vehicles (buses) can exchange information with the bus stop (infrastructure) which are connected to the backhaul network through communication technologies such as 5G or WiMAX (Worldwide Interoperability for Microwave Access).

6.4 Mobility Traces from SUMO Simulations

SUMO is a microscopic multi-modal traffic simulator which allows the behavior of each vehicle to be explicitly controlled and monitored [108]. SUMO has been extensively employed in various projects related to network performance, traffic assignment, vehicle routing, traffic impact analysis, V2X communication analysis, etc. In this section, the details of the simulation carried out in SUMO to obtain the synthetic mobility traces of the buses are discussed. The simulation is divided into two cases as described in Sections 6.4.1 and 6.4.2, where the required input parameters are set based on the details discussed in the respective sections.

SUMO provides an option to use on-board wireless communication device (usually bluetooth or wireless local area network) to analyze the nature of vehicular connectivity. Every vehicle can act as a sender or a receiver with configurable parameters like detection range and detection probability [116]. In our work we employ bluetooth as on-board wireless communication technology for the buses with a detection probability one, and detection range $R_v = 100$ m, i.e., given that two vehicles are within a communication range of 100 m, the probability that the

two vehicles identify each other is equal to one. For the V2I communication, the infrastructure is set up in SUMO by parking a vehicle at a designated position, and by enabling the communication capability (R_I=100 m) of the vehicle. The details of the simulation set-up are discussed below:

1. A simple real-world road network topology is imported from OpenStreetMap [109]. The imported road network topology includes # of lanes, junctions, traffic lights, POIs, bus stops and pedestrian walkways.

2. A route file describing the route details of individual bus routes including their time headway (inter-arrival time) and dwell time are provided as an input to the simulation. The # of routes and the dwell time for each route are determined by the two topological metrics k_i and w_i.

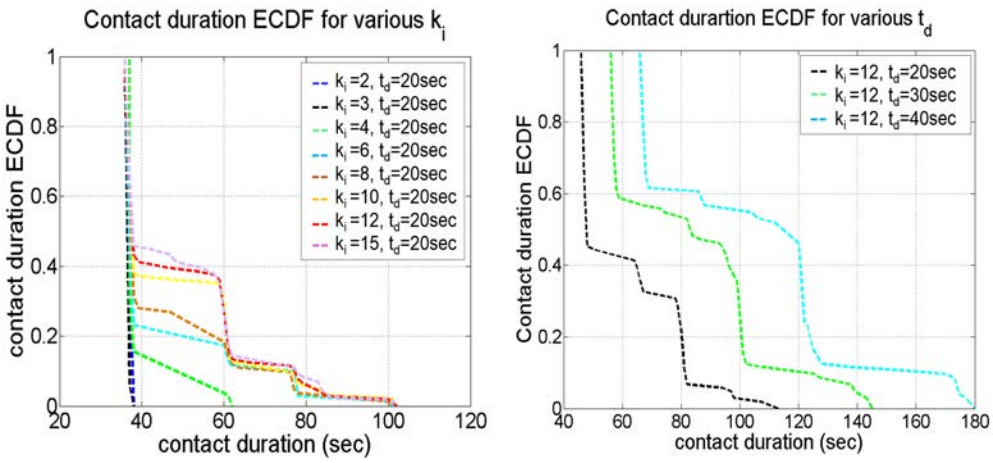
3. The simulation is run for the morning peak-hour for a duration of 3600 sec (8:00 - 9:00 AM) for different values of k_i and w_i.

4. At the end of the simulation, SUMO generates various trace files which captures the footprints of individual vehicle with a sampling period of 1 sec that includes details on the speed, wait time, geographic position of the bus (latitude-longitude), the lane ID, stop location, dwell time, start and end time of vehicle detection, etc. The two time stamps, t_{start}, t_{end} that fall within the range R_I are directly recorded in the footprints of SUMO which are employed to determine the V2I contact duration as given by (6.4).

In the next section, we discuss further details of the simulation by varying the two topological metrics, k_i and w_i (λ_i^m and t_d), which in turn determines the nature of V2I contact duration.

6.4.1 Case1: vary the weighted in-degree (k_i) of a node

In this case, to understand the V2I connectivity, we vary the metric k_i, and w_i is kept constant. The weighted in-degree of a node along with the time headway determines the density of buses around a bus stop. For example, let $k_i = 5$, then, $\lambda_i = \left\{\lambda_i^1, \lambda_i^2, .., \lambda_i^5\right\}$ such that $4 \leq \lambda_i^m \leq 10$, i.e., the time headway between the buses of same route is randomly assigned a value between 4 to 10 min to suit the real-world scenario for the peak hour simulation. Furthermore, since the node



(a) **(b)**

Fig. 6.3: Empirical CDF indicating the (a) variation of t_c with k_i; (b) variation of t_c with w_i or t_d.

weight is assumed to be constant, i.e., the number of people accessing a bus stop is constant, the dwell time for the buses is set to be a constant value of 20 sec.

By setting the parameter values as discussed above, the simulation is run, and the footprints are collected. Using the footprints, the V2I contact duration is evaluated. Fig 6.3a indicates the variation of contact duration for different weighted in-degree and constant dwell time. We observe a transition point called the *characteristic time* at $t_c \approx 40$ s above which the curves decay slower. Our results accords with the results observed by both [89, 91]. Also, as observed from the figure, the slope of the curves are dependent on the degree, i.e., higher the degree, slower the decay rate of contact duration, and thus a better connectivity.

Intuitively, this is an expected behavior since the density of the buses increases with increasing k_i which contributes towards increased t_c. Thus, based on our simulation results, we set the first condition to observe an increased contact duration around a bus stop, i.e., $k_i \rightarrow (k_i)_{\max}$.

6.4.2 Case2: vary the demand centrality (w_i) of a node

In this section, we vary the demand centrality of node by keeping the degree constant. As discussed in Section 3.3.1, the metric demand centrality is determined by NOP, which helps to determine the dwell time of the buses at a stop. Similar to Case 1, λ_i^m is randomly assigned a value such that $4 \leq \lambda_i^m \leq 10$. For simplicity,

we assume that, given the demand centrality of a node is high, irrespective of its degree, all the routes servicing the stop encounters equal demand. Furthermore, since we already observed in Section 6.4.1 that t_c improves when $k_i \to (k_i)_{\max}$, in this case we choose a high value of k_i to observe the simulation results.

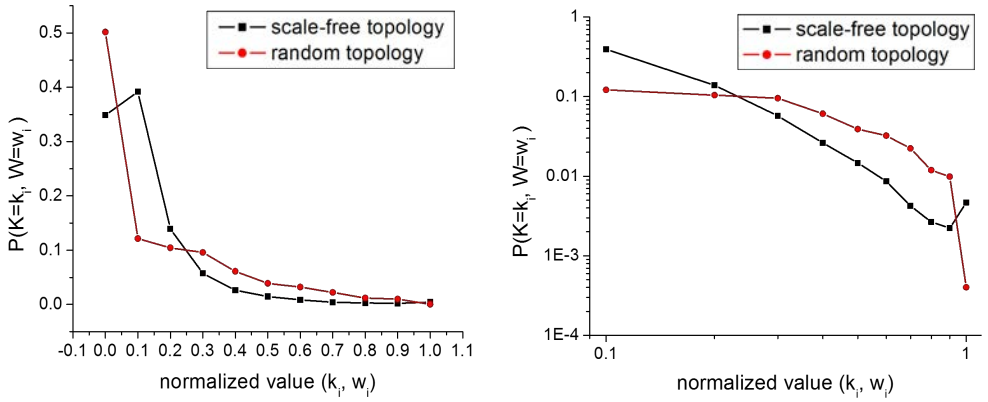
By setting the parameter values as discussed above, the simulation is run, and the footprints are collected. Using the footprints, the V2I contact duration is evaluated. Fig 6.3b indicates the variation of contact duration for different dwell time and constant degree. From figure, it is observed that, higher the dwell time, greater is the contact duration. Intuitively, this is again an expected behavior since the increased wait time of the buses contribute towards higher contact duration. However, with increased dwell time, the buses occupies the berth for a longer time, leading to less density of buses within a range R_I.

By comparing Figs. 6.3a and 6.3b, it can be concluded that the characteristic time is clearly dependent on the dwell time of the buses, whereas, the slope of the plots is dependent on the degree. Our simulation results accords with the results discussed in [89]. Thus, based on our simulation results, we set the first condition to observe an increased contact duration around a bus stop, i.e., $w_i \to (w_i)_{\max}$ (or $t_d \to (t_d)_{\max}$).

From the two simulation cases discussed in Sections 6.4.1 and 6.4.2, it can be stated that, as $k_i \to (k_i)_{\max}$ and $w_i \to (w_i)_{\max}$, then $t_c \to (t_c)_{\max}$. Thus, in the future sections we emphasize on the condition $(k_{i\max}, w_{i\max})$ which provides us the information about $(t_c)_{\max}$. Lastly, as part of microscopic connectivity analysis, the synthetic mobility traces from SUMO simulations were employed to demonstrate the influence of two topological metrics on the nature of V2I connectivity. The results of microscopic connectivity analysis are then used to study the V2I connectivity at a macroscopic level, as discussed in next section.

6.5 Vehicular Network Connectivity: A Macroscopic Analysis

For a given transport network topology, by studying the distribution pattern of the two topological metrics, we can determine the distribution pattern of the V2I contact duration (t_c), the details of which are discussed in this section. We consider two topologies: random and scale-free, to demonstrate the distribution pattern of V2I contact duration.



(a) **(b)**

Fig. 6.4: Joint PDF of the two topological metrics k_i and w_i in (a) linear; and (b) log-log scale for the two network topologies. In the figure, the scale-free topology represents the Hong Kong BTN, and random topology represents London BTN.

Now, let K and W be two random variables such that $K = k_i$ and $W = w_i$ $\forall i = 1, 2, .., N$, and $0 \leq k_i, w_i \leq 1$. For the ease of representation, the values of k_i and w_i are normalized such that the two metrics take the values 0, 0.1, 0.2, 0.3, ..., 0.9, and 1. Let the probability distribution function of the two topological metrics be denoted by $P(K = k_i)$ and $P(W = w_i)$ that are obtained from the results discussed in Chapter 3 and 4.

Using the distribution of individual random variable, we evaluate the conditional probability distribution function (PDF), i.e.,

$$\mathrm{P_{K,W}(k_i, w_i) = P(W = w_i | K = k_i) P(K = k_i); \quad k_i = w_i = 0, 0.1, 0.2, ..., 1} \qquad (6.5)$$

For a given value of k_i, though w_i can take values between 0 and 1, we are specifically interested in knowing the conditional joint PDF of the two random variables, since we already demonstrated the nature of linear dependency between k_i and w_i in Section 3.3.1 in Chapter 3, i.e., $P_{K,W}(k_i, w_i) = P(K = k_i \cap W = w_i)$.

Figs. 6.4a and 6.4b plot the joint PDF of the two random variables on a linear- and log-log scale, respectively. In the two figures: the low values of k_i and w_i (≤ 0.3) indicate that a poor t_c is observed around such nodes; the values of k_i and w_i between 0.3 to 0.9 indicate that a mediocre level of t_c is observed around such nodes; and the high values of k_i and w_i (≥ 0.9) indicate that a better t_c is observed around such nodes. However, as observed from the figure:

1. The nodes with low values of k_i and w_i (≤ 0.3) have higher probability of occurrence in a scale-free network topology. Thus, a poor V2I contact duration is expected to occur in high probability in a scale-free network topology.

2. The nodes with mediocre values of k_i and w_i (0.3 to 0.9) occur with high probability in random network topology. Thus, a mediocre nature of V2I contact duration is expected to occur in high probability in random network topology.

3. The nodes with high values of k_i and w_i (≥ 0.9) occur with high probability in scale-free network topology as compared to random network topology. Thus, in scale-free topology though a high V2I contact duration is expected, the probability of finding such nodes is less.

Thus, the two network topologies demonstrate significantly different distribution pattern of V2I contact duration. Specifically, the V2I contact duration for the two well-known topological structures: scale-free and random clearly demonstrates the power-law and Poisson distribution, respectively.

6.6 Topological Vs. Spatial Connectivity

The transport networks belong to the category of spatial networks. Hence, we further investigate the nature of V2I connectivity considering the spatial embedding of the two network topologies. To achieve this, we make use of supernodes that are existing in the network. We believe that a supernode can aid in further improving the V2I contact duration. Such spatial zones where the supernodes exist in a network can be used by the network operators to exchange the information that require higher throughput and longer contact duration.

The definition of a supernode as defined in (3.3) in Section 3.2 in Chapter 3 is a reasonable consideration from a transport network analysis, however, the definition might not be valid from a connectivity analysis perspective. That is, as demonstrated in Fig. 6.5b, the members of a supernode can be present on the same road segment (suitable for both BTN and V2I analysis), can be on either sides of a road segment (suitable for BTN analysis), or can be on the parallel road

(a) **(b)**

Fig. 6.5: An example of geographically closer nodes in Hong Kong which are termed as supernodes; (b) the supernode considered in BTN analysis and V2I connectivity analysis.

segment, near junctions, etc. (suitable for BTN analysis). For V2I connectivity analysis, a supernode whose member nodes lie on the same road segment are better suited, and thus the condition for defining a supernode is modified as

$$(d_{ij} \leq d_{th}\ )|\ \ i, j \in R';\ \ d_{th} = 100\ \text{m} \tag{6.6}$$

where d_{ij} is the geographic distance between two nodes i and j, and R' is a chosen bus route. Thats is, in VANET connectivity analysis, a supernode is a set of geographically closer nodes which satisfy the condition $d_{ij} \leq d_{th}$ such that the nodes i and j are along a specific bus route R'. Thus, the revised supernodes consists of nodes which typically lie on the same road segment, and are located within a range of 100 m.

Fig. 6.6 shows the variation of t_c around a regular node and a supernode, with an optimal value of k_i and w_i (or t_d). As observed from the figure, an improved value of t_c is observed around a supernode as compared to a regular node. Furthermore, the simulation results show that the behavior of t_c for a regular node (with optimal value of k_i and w_i) demonstrates an exponential distribution, whereas, for the supernode, demonstrates a power law distribution. For the purpose of V2I information exchange, a high value of t_c is expected, however, the power law nature of t_c might not be of much interest since it demonstrates the reduced probability occurrence of high value of t_c. A careful observation into Fig. 6.6 reveals the fact that $\approx 80\%$ of the entire distribution can be attributed to exponential nature, while the remaining 20% can be attributed to power law nature, which accords with the results discussed in ref. [91].

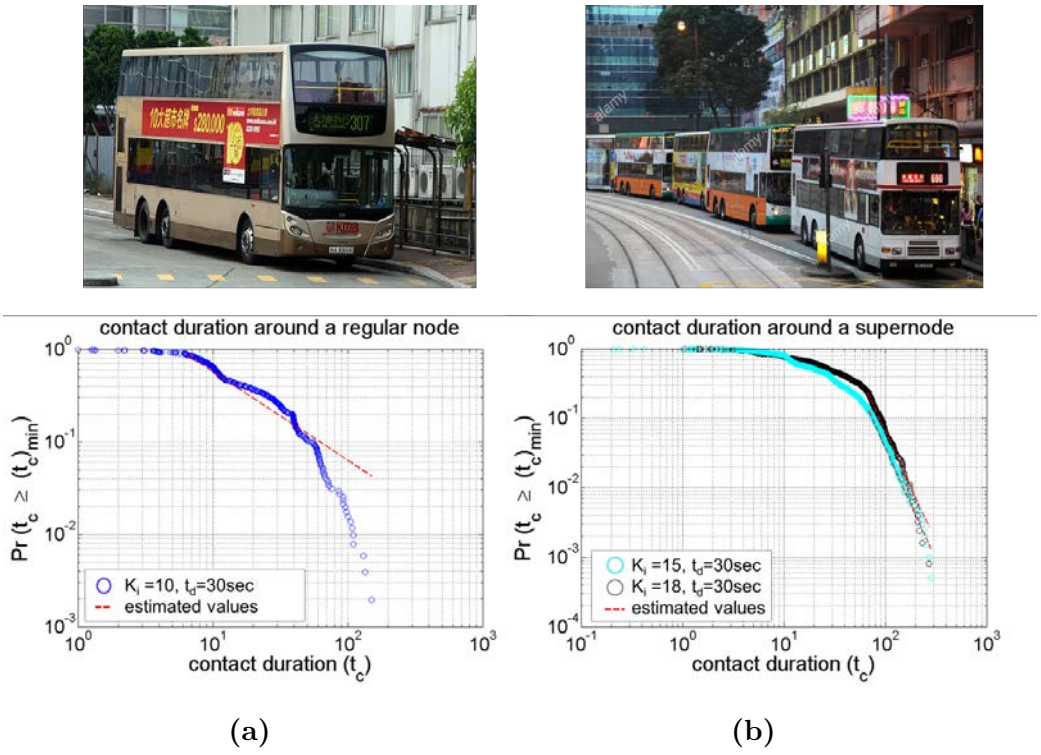


Fig. 6.6: The V2I connectivity pattern around a regular node and a supernode.

Thus, the results in Fig. 6.6 illustrate that the supernodes existing in the network can be used to improve the V2I contact duration. To demonstrate the probability of finding such nodes in vehicular network connectivity, we consider the third random variable i, along with K and W. The parameter i indicates whether a node is a regular node or a supernode. Considering the three random variables, we evaluate the conditional PDF as,

$$\mathrm{P_{K,W,i}(k_i, w_i, sn_i) = P(sn_i|k_i \cap w_i)P(w_i|k_i)P(k_i);\ \ k_i = w_i = 0, 0.1, 0.2, .., 1} \quad (6.7)$$

Fig. 6.7a shows the conditional PDF of the three random variables for a specific case for the two network topologies: scale-free (Hong Kong) and random (London). That is, the figure shows the probability of finding the supernodes in a network for a chosen value of k_i and w_i. The specific scenario is observed since it satisfies the condition for obtaining the maximum contact where, $k_i \rightarrow (k_i)_{\max}, w_i \rightarrow (w_i)_{\max}$, and a node i behaves as a supernode.

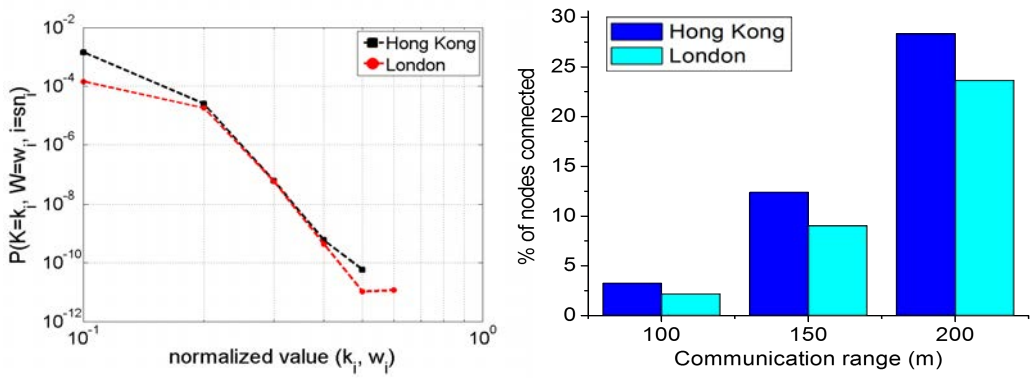
As observed from the figure, with increasing values of k_i and w_i, the probability of finding a supernode decreases drastically. From a transport network analysis perspective, this is an expected behavior where the nodes with high degree and

high demand are seldom deployed geographically closer. However, the probability of finding a supernode for low values of k_i and w_i is quite high. Such a behavior of nodes can be exploited by the communication engineers to improve the nature of connectivity among nodes with low degree and demand, which are found in higher probability in both the random and scale-free network topologies.

The impact of topological analysis on VANET connectivity in Section 6.5 revealed that in both scale-free and random network topologies the probability of observing poor V2I connectivity is higher, however, by understanding the spatial embedding of the topology we further revealed that such poor connectivity can be slightly improved by exploiting the supernodes available in a network. Thus, to assess the impact of underlying transport network topology on the vehicular network connectivity, merely considering the topological analysis without considering its spatial embedding might fail to capture the true nature of connectivity, and thus demands for such a consideration in the future works.

Finally, in defining a supernode, by varying the parameter d_{th}, we can observe the inherent connectivity among the infrastructure (bus stops), i.e., along with observing the V2I connectivity around the bus stop within a range R_I, we can also observe the I2I (infrastructure-to-infrastructure) connectivity. Fig. 6.7b indicates the histogram showing the percentage of bus stops that are connected among themselves for various vales of d_{th} or R_I for the two network topologies. As observed from the figure, with a small communication range of $R = 200m, \approx 25\%$ of the infrastructure is effortlessly connected. Of course, considering the spatial embedding of bus stops, relay devices can be employed at appropriate locations to further enhance the connectivity of the network, both V2I and I2I. Such special zones where the bus stops are deployed closer can be used to develop robust and more efficient routing algorithms since a more sustained and improved contact duration is expected.

Also, we assumed in the beginning that all the bus stops in the network are connected to the backhaul through communication technologies such as 5G or LTE (Long Term Evolution), however, connecting the entire network to the backhaul using 5G can prove to be expensive. By understanding the I2I connectivity, we can judiciously choose the set of bus stops that can be connected among themselves in small-scale using less expensive technologies such as WLAN or DSRC,



(a) **(b)**

Fig. 6.7: (a) The conditional PDF of the three random variables; (b) the inherent connectivity of infrastructure for different communication ranges.

on the other hand we can choose the set of bus stops that can be connected to the backhaul using expensive technologies such as 5G. Such connectivity leads to a layered vehicular connectivity network that can greatly assists the communication engineers in utilizing the existing infrastructure to offer seamlessly connectivity to both the vehicles and passengers using it. In the next section, we verify the robustness of the two VANET communication topologies: random and scale-free, to demonstrate the capability of the two topologies to offer the best possible connectivity while some of its nodes fail to function.

6.7 Vehicular Network Connectivity Robustness Analysis

Many real-world complex systems have demonstrated an exceptional ability to retain their basic functionality even in the case of failure of their network components, termed as the network robustness. The primary interest in robustness analysis is to understand the impact of nodes failure on the integrity of a network. In this section, we discuss some fundamental results on the topological robustness of the vehicular network connectivity by employing the random node removal approach.

To assess the topological robustness, the commonly employed approach is to randomly remove a set of nodes from the network which quantify high value of degree- and betweenness centralities, whereas, in our work, we consider a more contemporary definition to assess the communication capability of a network

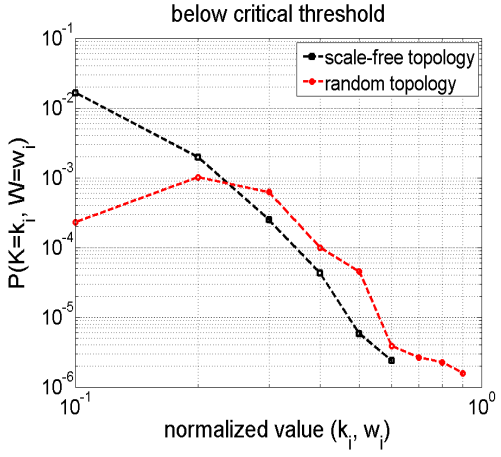
under node failure condition. By random node removal, we assume that a node is not capable of exchanging V2I information. Thus, from the VANET connectivity analysis perspective, the removal of a node or failure in its functionality will not have any effect on its associated edges.

Considering this scenario, we remove a set of nodes from the two network topologies randomly to indicate that the nodes are no more capable of performing their normal functionality (i.e., V2I communication). Later, we evaluate the joint PDF $P(K = k_i \cap W = w_i)$ as given by (6.5) for the two networks with the remaining set of nodes. This step is carried out multiple times until we reach the critical threshold for removing the nodes. A detailed discussion regarding the critical threshold for the two topologies: random and scale-free is discussed by [5, Chapter 8]. Fig. 6.8 shows the nature of V2I connectivity for the two network topologies under three scenarios, i.e., number of node failures below the critical threshold, at the critical threshold, and above the critical threshold.

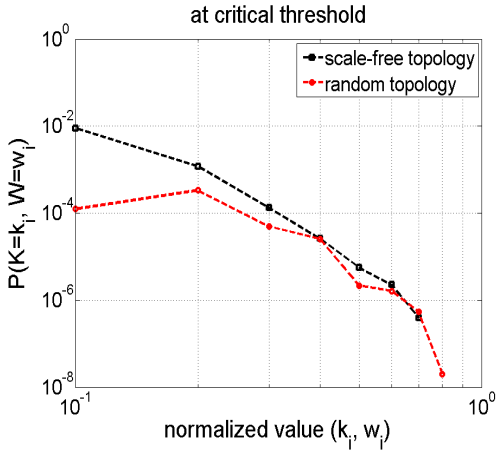
As discussed in Section 6.5, for a large-scale information dissemination in PTNs, the random network topology offers a more dispersed and optimal V2I connectivity, as compared to more concentrated and poorer V2I connectivity in scale-free networks. Whereas, from Fig. 6.8, we can observe that, under the random node failure condition, the scale-free topology, though offers poorer connectivity, has a sustained connectivity offering the best possible service even under critical conditions. Whereas the random network topology demonstrates a significant degradation in the service as the number of node failures increase.

6.8 Summary

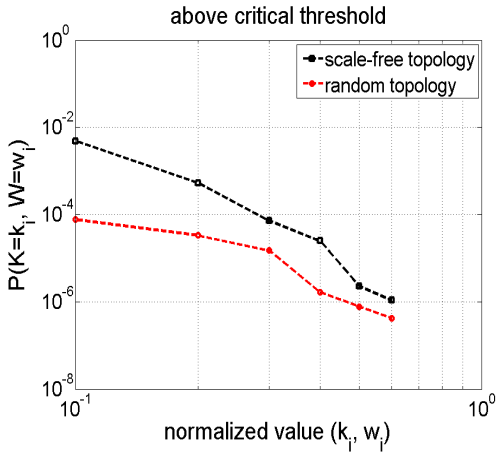
In the first of its kind, we aimed at integrating the two different disciplines of study to demonstrate the impact of transport network topology on vehicular network connectivity. Though the work carried out in the chapter is a high-level analysis, such an analysis provides some fundamental results for future work to consider the impact of underlying topological behavior in assessing the dynamics of vehicular network connectivity. Our work in this chapter provides insight into the large-scale information dissemination of non-time-critical information like infotainment applications, broadcasting advertisements on the buses, relaying touristic information, social media updates, etc.



(a)



(b)



(c)

Fig. 6.8: The joint PDF of k_i and w_i for the two topologies under random node removal approach (a) below critical threshold; (b) at critical threshold; (c) above critical threshold. In the figure, the scale-free topology represents the Hong Kong BTN, and random topology represents the London BTN.

For the part of microscopic connectivity analysis, we considered two topological metrics: wighted in-degree and demand centrality of a node, to understand the nature of V2I contact duration around a node, within a range R_I. The results from microscopic analysis were scaled-up to macroscopic analysis by considering the distribution of the two topological metrics for a given network topology. The results of macroscopic analysis showed that the random network topology comprised of nodes which offered an optimal V2I contact duration in high probability, whereas, the scale-free topology comprised of a small set of nodes that offered high V2I contact duration while other set of nodes offered poor V2I contact duration. Thus, depending on the level of information dissemination, the two topologies can be employed for different applications in VANETs since they demonstrate two significantly different behaviors.

Furthermore, we also demonstrated that supernodes existing in the network can be exploited by the communication engineers to improve the V2I contact duration. Also, by considering the spatial embedding of a transport network topology, the special zones such as supernodes can be used by the communication engineers to develop more robust and realistic protocols for information exchange in VANETs.

By using the approach of random node failure condition, we demonstrated that though the scale-free topology is more robust as compared to random topology. That is, the scale-free topology though offered poor connectivity, it showed a sustained connectivity under critical conditions, unlike the random network topology. Such an analysis provides great insight to the network deployers on the capability of the communication network to sustain during malfunctions, and yet offer the best possible service.

Chapter 7

Conclusions and Future Work

In this chapter, we re-iterate the main contributions of the thesis and discuss some potential topics for future research.

7.1 Main Contributions of the Thesis

The main contributions of the thesis include:

1. Representation of public transport network topology as a directed and weighted graph in supernode representation.

   To accomplish the PTN topological analysis, based on various statistical information, we chose three cities: Hong Kong, London, and Bengaluru. Our study was limited to two major public transport modes: bus and metro modes. For the analysis of bus transport networks, a directed and weighted graph structure was employed, whereas, an undirected and weighted graph structure was employed for metro transport analysis. The main reason for considering the graph type is the level of overlapping among the inbound and outbound routes. The L-space representation was chosen since it closely mimics the real-world infrastructure of a PTN.

   Considering the spatial embedding of PTNs, a new network element called supernode was defined, and a novel graph representation called supernode graph representation was proposed, to model the PTN topology. Static weights were added to both nodes and edges to effectively capture their significance in the PTN analysis since the consideration of real-world parameters as weights aids in improved understanding of the network behavior. Finally, the proposed zone classification approach alongside with the

node weight approach demonstrated the effective way of capturing a node's significance to the network.

2. Topological analysis of bus transport network from a graph theoretic perspective: a mono-layer analysis.

   Topological analysis of the three bus transport structures was accomplished using various local metrics like degree, clustering, betweenness centrality, closeness centrality, and global metrics like degree distribution, scale-free property, average path length, and small-world property. The study of various local and global properties provided intriguing information about the topological behavior of BTNs. Such a study can be a great source of information for researchers in the applied fields such as designing of transfer algorithm, optimization of public transport routes, network planning, and transit operation. Additionally, we illustrated at various instances that the supernode structure offered practical and more insightful perspective in understanding the true network behavior which was difficult to be captured by the conventional graph representation. The validation of topological efficiency for a specific route using the results from SUMO simulation and real-world data from KMB demonstrated an effective way to measure the time efficiency of the topology, and also illustrated the dependency of vehicular mobility on the node weight.

3. Topological analysis of bus and metro transport networks from a graph theoretic perspective: a multi-layer analysis

   A simple approach of spatial amalgamation was proposed to capture the interdependency between the two transport layers: bus and metro networks. Node weight approach was employed to identify a set of influential nodes by considering and ignoring the layer interdependency. The ranking of nodes using node weight approach demonstrated that ignoring the interdependency between the two transport modes failed to capture the true significance of a node in view of practical relevance. Thus, using the approach of spatial amalgamation, we demonstrated that, to simultaneously capture the benefaction of the nodes and edges to multiple layers, modeling the PTN topology as an interdependent network is necessary.

4. Impact of PTN topology on the vehicular network connectivity.

   For microscopic analysis, two topological metrics: wighted in-degree and the demand centrality of a node were considered to understand the nature of V2I contact duration. For macroscopic analysis, the distribution pattern of the two topological metrics for a given network topology was studied to determine the distribution pattern of V2I contact duration. The results from the macroscopic analysis revealed that for a large scale information dissemination in VANETs, the random network topology offered better and more dispersed V2I contact duration as compared to the scale-free topology. However, considering the spatial embedding of the topology, the supernodes existing in the network can be exploited to further improve the nature of V2I contact duration. The robustness of the vehicular network was assessed using the random node removal approach which demonstrated that the scale-free topology offered poor yet sustained connectivity, thus offering the best possible service under critical conditions, unlike the random network topology which demonstrated significant degradation in the service.

Finally, some of the key take away points from our work are: (a) applying the concepts of graph theory to the analysis of public transport networks offered an effective and convenient way to understand the network operation at both local and global levels, and from both operator and passenger perspectives; (b) considering supernodes in a PTN analysis not only rendered useful information on topological behavior, but also benefited in exploring the hidden network behaviors which was difficult to be captured with conventional graph representation; (c) in view of the practical relevance of the network, we demonstrated that ignoring the interdependency among multiple transport modes can be expensive at the cost of assessing the benefaction of nodes and edges to multiple layers; and (d) the impact of topological analysis in analyzing the dynamics of vehicular network connectivity demonstrated that the topological metrics and their distribution pattern should be taken into consideration by communication engineers in developing robust and realistic information dissemination approaches.

7.2 Scope for Future Work

In a data driven world, the availability of real-world datasets and the high-end tools for handling huge volume of data have greatly facilitated the research of complex systems and data analysis. Despite the successful attempts in applying the concepts from network science to PTN analysis, rigorous study of PTN from a network science perspective is still relatively rare. For example, while PTN topological 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 that is desirable at the operational level.

Alongside with offering a few advantages, the network-based analysis also raises a few technical challenges as a consequence of increased computational time with increasing network size which should be addressed in the future. Furthermore, while research efforts have been devoted to the spatial analysis of PTNs, the temporal dynamics reflecting the topological variation of a PTN deserves immediate attention. Adding to the spatio-temporal representation of a PTN topology, consideration of dynamic connectivity, and the study of distributed passenger flow from a global view point should attract more research interest to capture the dynamic topological behaviors which is close to the real-world scenario. For example, the mechanism of passenger flow in a scale-free network might be significantly different as compared to random network topology.

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. Identifying the supernodes which behaves as the actual passenger interchange

stations within and between the transport modes should be considered with respect to the real-world network usage. This can offer insightful suggestions to the operators to design holistic transport systems.

Another core research area of practical importance is robustness analysis which aims to study the network functionality upon failure of certain nodes or edges. Combining the study of robustness analysis and the passenger flow mechanism in an integrated multi-layer network, insightful results can be gathered by network engineers pertaining to the capacity of a network in offering the best possible service during emergency failure in its components, which is one of the most prevalent scenarios in the PTNs.

Lastly, in the future, the public transport stops are not only merely expected to be points of passenger carriers, but are also expected to be the information carriers such that any information update from the network operators could reach a significant amount of passengers accessing the PTNs within the least possible amount of time. The existing infrastructure of the PTNs can be exploited by the network deployers to prepare them as future smart infrastructures to offer better connectivity to both the on-road vehicles and passengers, and thus contribute towards developing the future of smart and connected cities a reality.

APPENDIX

A 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 B_{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	-

B Online Sources

	Source	Ref.
	Bus transport network	
Hangzhou	www.hzbuda.com.cn	[24]
Chennai	www.mtcbus.org/	[17]
Ahmadabad	www.ahmedabadbrts.org/web/commuters.html	[17]
Delhi	delhitravelhelp.in/StopsOfBus.aspx	[17]
Hyderabad	www.hyderabadbusroutes.com	[17]
Kolkata	www.kolkataonline.in	[17]
Mumbai	github.com/transitmetrics/ntd/tree/master	[17]
Singapore	www.streetdirectory.com.sg/	-
Hong Kong	data.gov.hk/en-data/category/transport?organization= hk-td/	[117]
London	data.london.gov.uk/dataset/tfl-bus-stop-locations-and-routes	-
Bengaluru	opencity.in/topic/transportation/	-
Australia	opendata.transport.nsw.gov.au/search/type/dataset	-
-	www.apta.com	[14]
	Metro transport network	
Beijing	www.ebeijing.gov.cn/feature_2/BeijingSubway/	[34]
Shanghai	service.shmetro.com/en/	[32]
Hong Kong	www.mtr.com.hk/en/customer/tourist/index.php	[33]
Tokyo	www.tokyometro.jp/en/subwaymap/index.html	[33]
London	tfl.gov.uk/maps/track?intcmp=40400	[33]
New York	web.mta.info/maps/submap.html	[33]
Boston	mbta.com/schedules/subway	[118]
Paris	parisbytrain.com/paris-metro/	[33]
Seoul	www.korea4expats.com/korean-subways.php	[36]

C KMB Dataset

Below is an excerpt of the information provided by one of the major bus operator in Hong Kong, the Kowloon Motor Bus Co. (KMB). The data was collected for a period of one week (22/10/2017 to 28/10/2017) for morning and evening rush hours (8:00 - 11:00AM, 3:00 - 7:00PM) for a round-trip bus route (#1 Chuk Yuen Estate Bus Terminus to Tsim Sha Tsui Star Ferry). The information on number of passenger alighting at a stop is lacking in the provided dataset since the exit count of passengers is not monitored by the operators.

bus_no	stop_name_chi	stop_name_EN	bound	PAX_count	arrival_time
RJ5466	竹園⌧總站	Chuk Yuen Estate terminus	1	28	7:47:00 AM
RJ5466	天虹小學	Rainbow primary school	1	7	7:48:15 AM
RJ5466	馬仔坑遊樂場	Ma Son pit playground	1	1	7:50:20 AM
RJ5466	摩士公園	Morse Park	1	1	7:52:19 AM
RJ5466	摩士公園體育館	Morse Park Stadium	1	0	7:53:05 AM
RJ5466	偉東樓	Wei East House	1	6	7:54:09 AM
RJ5466	盈東樓	Ying East Building	1	3	7:54:57 AM
RJ5466	九龍寨城公園	Kowloon Walled City Park	1	3	7:56:27 AM
RJ5466	賈炳達道	Carpenter Road	1	5	7:58:46 AM
RJ5466	太子道西	Prince Edward Road West	1	0	8:01:05 AM
RJ5466	喇沙利道	La Salle Lane	1	0	8:03:37 AM
RJ5466	伯爵街	Earl Street	1	2	8:04:25 AM
RJ5466	勵德街	Reed Street	1	0	8:05:19 AM
RJ5466	拔萃男書院	Diocesan boys college	1	0	8:06:03 AM
RJ5466	協和小學	Concord Primary School	1	0	8:07:27 AM
RJ5466	太子站	Prince station	1	0	8:08:07 AM
RJ5466	旺角奶路臣街	Mong Kok Nelson Street	1	0	8:10:45 AM
RJ5466	旺角豉油街	Mong Kok Soy Street	1	0	8:11:46 AM
RJ5466	油麻地永星里	Yau Ma Tei Wing Star	1	3	8:14:22 AM
RJ5466	油麻地長樂街	Yau Ma Tei Changle Street	1	0	8:16:13 AM
RJ5466	尖沙咀德成街	Tak Sing Street, Tsim Sha Tsui	1	0	8:18:06 AM
RJ5466	金馬倫道	Cameron Road	1	0	8:20:04 AM
RJ5466	尖沙咀中間道	Tsim Sha Tsui Middle Road	1	0	8:21:38 AM
RJ5466	香港文化中心	with love	1	0	8:23:50 AM
RJ5466	尖沙咀碼頭	Tsim Sha Tsui Ferry Pier	1	NULL	NULL

(a)

bus_no	Log_datetime	speed	gps_pos_lat	gps_pos_long
RJ5466	3:44:32 PM	0	22.340007667	114.154614000
RJ5466	3:44:33 PM	0	22.340007667	114.154614000
RJ5466	3:44:34 PM	0	22.340007667	114.154614000
RJ5466	3:44:35 PM	0	22.340007667	114.154614000
RJ5466	3:44:36 PM	0	22.340007667	114.154614000
RJ5466	3:44:37 PM	0	22.340007667	114.154614000
RJ5466	3:44:38 PM	0	22.340007667	114.154614000
RJ5466	3:44:39 PM	0	22.340007667	114.154614000
RJ5466	3:44:40 PM	0	22.340007667	114.154614000
RJ5466	3:44:41 PM	0	22.340007667	114.154614000
RJ5466	3:44:42 PM	0	22.340007667	114.154614000
RJ5466	3:44:43 PM	0	22.340007667	114.154614000
RJ5466	3:44:44 PM	0	22.340007667	114.154614000
RJ5466	3:44:45 PM	0.4	22.340007667	114.154614000
RJ5466	3:44:46 PM	2.7	22.340011667	114.154606333
RJ5466	3:44:47 PM	4.9	22.340011333	114.154602167
RJ5466	3:44:48 PM	6.7	22.340003000	114.154605000
RJ5466	3:44:49 PM	9.9	22.339993000	114.154611667
RJ5466	3:44:50 PM	10.3	22.339983833	114.154627000
RJ5466	3:44:51 PM	10.3	22.339973333	114.154643333
RJ5466	3:44:52 PM	9.4	22.339962167	114.154660000
RJ5466	3:44:53 PM	9.4	22.339951833	114.154676333
RJ5466	3:44:54 PM	6.3	22.339941833	114.154688500
RJ5466	3:44:55 PM	2.2	22.339937500	114.154693000
RJ5466	3:44:56 PM	0	22.339938000	114.154690500
RJ5466	3:44:57 PM	0	22.339938000	114.154690500
RJ5466	3:44:58 PM	0	22.339938000	114.154690500

(b)

Bibliography

[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," *NASA Report CR-2005-213522*, 2005.

[2] A.-L. Barabási, "Linked: The new science of networks," 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 19th Century Germany*. University of California, Berkeley, 2008.

[5] A.-L. Barabási, *Network science*. Cambridge university 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] C. Sommer and F. Dressler, *Vehicular networking*. Cambridge University Press, 2014.

[9] E. CLIFF, "Lewis mumford and norman bel geddes: the highway, the city and the future," *Planning Perspectives*, vol. 20, no. 1, pp. 51–68, 2005.

[10] S. Takaba, "Japanese projects on automobile information and communication systems-things aimed at and obtained in 20 years' experiences," in *Vehicle Navigation and Information Systems Conference*, vol. 2, pp. 233–240, IEEE, 1991.

[11] "Prometheus Project." http://www.eurekanetwork.org/project/id/45. [Online; accessed 10-Aug-2018].

[12] "LinkNY." https://www.link.nyc/. [Online; accessed 10-Jun-2018].

[13] M. Kurant and P. Thiran, "Extraction and analysis of traffic and topologies of transportation networks," *Physical Review E*, vol. 74, no. 3, pp. 036114(1)–036114(10), 2006.

[14] 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.

[15] 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.

[16] 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, pp. 1–7, 2013.

[17] A. Chatterjee, M. Manohar, and G. Ramadurai, "Statistical analysis of bus networks in india," *PloS one*, vol. 11, no. 12, pp. 0168478(1)–0168478(16), 2016.

[18] A. Háznagy, I. Fi, A. London, and T. Németh, "Complex network analysis of public transportation networks: A comprehensive study," in *International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 371–378, IEEE, 2015.

[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, pp. 1–9, 2017.

[22] 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.

[23] A. Chatterjee, "Studies on the structure and dynamics of urban bus networks in indian cities," *arXiv:1512.05909*, pp. 1–47, 2015.

[24] 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.

[25] 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.

[26] 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, pp. 1–13, 2016.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] S. Derrible, "Network centrality of metro systems," *PloS One*, vol. 7, no. 7, pp. 1–10, 2012.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] P. Angeloudis and D. Fisk, "Large subway systems as complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 553–558, 2006.

[38] P. Erdös and A. Rényi, "On random graphs i," *Publications Mathematicae*, vol. 6, pp. 290–297, 1959.

[39] P. Erdős and A. Rényi, "On the evolution of random graphs," 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] 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.

[44] G. Fagiolo, "Clustering in complex directed networks," *Physical Review E*, vol. 76, no. 2, pp. 026107(1)–026107(8), 2007.

[45] C. Von Ferber, Y. Holovatch, and V. Palchykov, "Scaling in public transport networks," *arXiv preprint cond-mat/0501296*, pp. 1–9, 2005.

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

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

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

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

[50] 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.

[51] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguná, G. Caldarelli, *et al.*, "Quantifying randomness in real networks," *Nature communications*, vol. 6, pp. 1–10, 2015.

[52] 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.

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

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

[55] 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.

[56] 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.

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

[58] 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.

[59] 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.

[60] M. E. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, pp. 208701(1)–208701(4), 2002.

[61] 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.

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

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

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

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

[66] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.

[67] M. Tomasini, "An introduction to multilayer networks," BioComplex Laboratory, Florida Institute of Technology, Melbourne, USA, pp. 1–14, 2015.

[68] M. Zanin, "Can we neglect the multi-layer structure of functional networks?," *Physica A: Statistical Mechanics and its Applications*, vol. 430, pp. 184–192, 2015.

[69] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," *Nature physics*, vol. 8, no. 1, pp. 40–48, 2012.

[70] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," *Nature communications*, vol. 6, pp. 6868(1)–6868(6), 2015.

[71] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, and G. Bianconi, "Weighted multiplex networks," *PloS one*, vol. 9, no. 6, pp. 1–8, 2014.

[72] P. S. Chodrow, Z. Al-Awwad, S. Jiang, and M. C. González, "Demand and congestion in multiplex transportation networks," *PloS one*, vol. 11, no. 9, pp. 1–10, 2016.

[73] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of a network of networks," *Physical Review Letters*, vol. 107, no. 19, pp. 195701(1)–195701(5), 2011.

[74] G. Baxter, S. Dorogovtsev, A. Goltsev, and J. Mendes, "Avalanche collapse of interdependent networks," *Physical review letters*, vol. 109, no. 24, pp. 248701(1)–248701(5), 2012.

[75] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, "The extreme vulnerability of interdependent spatially embedded networks," *Nature Physics*, vol. 9, no. 10, pp. 667–672, 2013.

[76] I. W.-H. Ho, K. K. Leung, and J. W. Polak, "Connectivity dynamics for vehicular ad-hoc networks in signalized road systems," in *21st International Teletraffic Congress, ITC 21 2009*, pp. 1–8, IEEE, 2009.

[77] I. W.-H. Ho, K. K. Leung, and J. W. Polak, "Stochastic model and connectivity dynamics for vanets in signalized road systems," *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 1, pp. 195–208, 2011.

[78] K. Rupinder and S. Gurpreet, "Survey of various mobility models in vanets," *International Journal Of Engineering And Computer Science*, vol. 3, no. 03, 2014.

[79] H. Boeglen, B. Hilt, P. Lorenz, J. Ledy, A.-M. Poussard, and R. Vauzelle, "A survey of v2v channel modeling for vanet simulations," in *Eighth International Conference on Wireless On-Demand Network Systems and Services*, pp. 117–123, IEEE, 2011.

[80] Z. Zhao, X. Cheng, M. Wen, B. Jiao, and C.-X. Wang, "Channel estimation schemes for ieee 802.11 p standard," *IEEE intelligent transportation systems magazine*, vol. 5, no. 4, pp. 38–49, 2013.

[81] F. D. Da Cunha, A. Boukerche, L. Villas, A. C. Viana, and A. A. Loureiro, *Data communication in VANETs: a survey, challenges and applications*. PhD thesis, INRIA Saclay, INRIA, 2014.

[82] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *IEEE Vehicular technology magazine*, vol. 2, no. 2, 2007.

[83] H. Feng and Y. Xu, "An empirical study on evolution of the connectivity for vanets based on taxi gps traces," *International Journal of Distributed Sensor Networks*, vol. 12, no. 2, pp. 2580465(1)–2580465(11), 2016.

[84] D. Naboulsi and M. Fiore, "Characterizing the instantaneous connectivity of large-scale urban vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1272–1286, 2017.

[85] L. Qiao, Y. Shi, and S. Chen, "An empirical study on the temporal structural characteristics of vanets on a taxi gps dataset," *IEEE Access*, vol. 5, pp. 722–731, 2017.

[86] N. Loulloudes, G. Pallis, and M. D. Dikaiakos, "The dynamics of vehicular networks in urban environments," *arXiv:1007.4106*, 2010.

[87] N. Loulloudes, G. Pallis, and M. D. Dikaiakos, "The dynamics of vehicular networks in large-scale urban environments," in *IEEE Conference on Collaboration and Internet Computing (CIC)*, pp. 192–199, IEEE, 2015.

[88] Y. Chen, M. Xu, Y. Gu, P. Li, L. Shi, and X. Xiao, "Empirical study on spatial and temporal features for vehicular wireless communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, pp. 1–12, 2014.

[89] I. W.-H. Ho, R. J. North, J. W. Polak, and K. K. Leung, "Effect of transport models on connectivity of interbus communication networks," *Journal of Intelligent Transportation Systems*, vol. 15, no. 3, pp. 161–178, 2011.

[90] M. Doering, W.-B. Pöttner, T. Pögel, and L. Wolf, "Impact of radio range on contact characteristics in bus-based delay tolerant networks," in *Eighth International Conference on Wireless On-Demand Network Systems and Services (WONS)*, pp. 195–202, IEEE, 2011.

[91] Y. Li, D. Jin, Z. Wang, L. Zeng, and S. Chen, "Exponential and power law distribution of contact duration in urban vehicular ad hoc networks," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 110–113, 2013.

[92] R. Albert, "Attack and error tolerance in complex networks," *Nature*, vol. 406, pp. 387–482, 2000.

[93] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks," *The European Physical Journal B*, vol. 71, no. 1, pp. 125–137, 2009.

[94] J.-L. Guillaume, M. Latapy, and C. Magnien, "Comparison of failures and attacks on random and scale-free networks," in *International Conference on Principles of Distributed Systems*, pp. 186–196, Springer, 2004.

[95] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Public transport networks under random failure and directed attack," *arXiv:1002.2300*, 2010.

[96] Z. Zou, Y. Xiao, and J. Gao, "Robustness analysis of urban transit network based on complex networks theory," *Kybernetes*, vol. 42, no. 3, pp. 383–399, 2013.

[97] S. F. Sousa, C. R. Neto, and F. F. Ferreira, "Structure and robustness of s\˜ ao paulo public transport network," *arXiv:1808.08117*, 2018.

[98] L. T. Authority *et al.*, "Passenger transport mode shares in world cities," *Journeys. Singapore: Land Transport Authority*, 2011.

[99] J. Lee and D. W. Wong, *Statistical analysis with ArcView GIS*. John Wiley & Sons, 2001.

[100] H. J. Miller and S.-L. Shaw, *Geographic information systems for transportation: principles and applications*. Oxford University Press on Demand, 2001.

[101] 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, pp. 1–9, 2017.

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

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

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

[105] "Bengaluru Dataset." http://opencity.in/topic/transportation/. [Online; accessed 16-April-2017].

[106] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[107] S. M. e. a. Ruowen Liu, Porter Beus, "Analysis of watts-strogatz networks," pp. 1–6, 2015.

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

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

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

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

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

[113] "Hong Kong Population Census Dataset." https://www.censtatd.gov.hk/hkstat/sub/sp130.jsp?productCode=FA100065. [Online; accessed 1-Feb-2016].

[114] "Travel Patterns London Dataset." https://data.london.gov.uk/dataset/travel-patterns-and-trends-london/. [Online; accessed 10-Jan-2017].

[115] C. G. Anand, "Presentation on bus-based public transport in bangalore bmtc experience," 2013.

[116] M. Behrisch and G. Gurczik, "Modelling bluetooth inquiry for sumo," in *Modeling Mobility with Open Data*, pp. 223–239, Springer, 2015.

[117] T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse, "Bus transport network in hong kong: Scale-free or not?," *Proc. of the International Symposium on Nonlinear Theory and Its Applications (NOLTA)*, pp. 610–613, 2016.

[118] 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.