Multi-layer Public Transport Network Analysis

Conference Paper · May 2018

DOI: 10.1109/ISCAS.2018.8351818

5 authors, including:

Tanuja Shanmukhappa
The Hong Kong Polytechnic University

Ivan Wang-Hei Ho
The Hong Kong Polytechnic University

Chi Kong Tse
City University of Hong Kong

Xingtang Wu
Beihang University (BUAA)

Multi-layer Public Transport Network Analysis

Tanuja Shanmukhappa[1], Ivan W. H. Ho[1,2], Chi K. Tse[1], Xingtang Wu[3] and Hairong Dong[3]
[1]Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
[2]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China
[3]Beijing Jiaotong University, Beijing, China
Email: tanuja.shanmukhappa@connect.polyu.hk, ivanwh.ho@polyu.edu.hk, encktse@polyu.edu.hk

***Abstract*—In this paper, we propose a novel method called supernode graph structure representation to model the public transport network structure of the London city. Supernode is a set of geographically closely associated nodes. Using the supernode graph structure, the bus transport and the metro transport network structures are analyzed by treating them as independent mono-layer or multi-layer network structures. A method of spatial amalgamation is proposed to integrate the two transport layers. A set of most influential nodes in the network is identified by assigning node weight to each node with respect to both mono-layer and multi-layer analysis. The behavior of these influential nodes is better characterized by categorizing them as either emitter, absorber or neutral zones.**

I. Introduction

Public transport networks (PTN) contain multiple layers of traffic carrying networks such as bus network, subway or metro network, tram network, ferry network, etc. Existing works in the analysis of PTNs using graph theory and complex networks have focused on the analysis of spatial or temporal network structure either by considering the individual transport layers or the overall transport layer (mapping the individual network layers as a single projected network layer). 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.

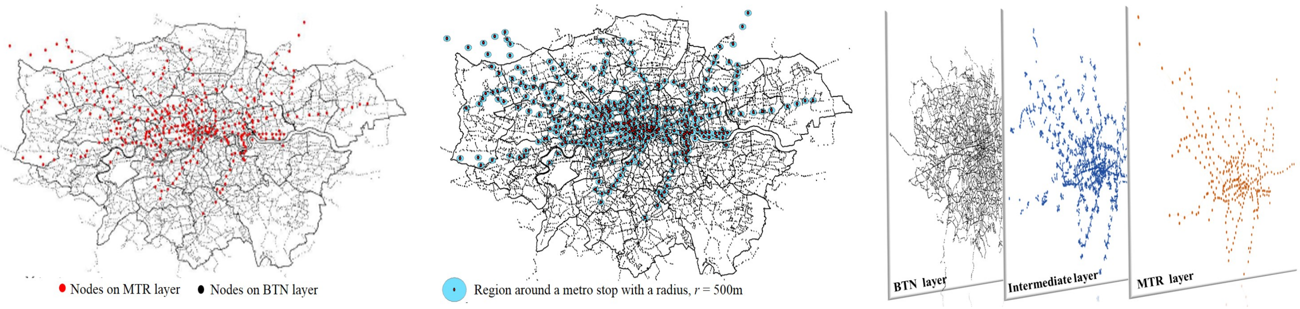
Kivela *et al.* [1] presented a detailed literature review by introducing a general framework for the multi-layer analysis of a wide range of networks. They provided exhaustive mathematical representation and notations for the monoplex (mono-layer) networks, multiplex (multi-layer networks), interdependent networks, interconnecting networks, networks of networks, etc. Tomasini [2] followed up on Kivela's work by introducing measures for multi-layer analysis in addition to the measures already introduced for mono-layer analysis in [1]. Zanin [3] discussed the multi-layer nature of the functional networks by rightly questioning the validity of a single type of edges endowed on nodes existing at different levels. The paper demonstrated that the results of the analysis of single projected layer might yield a biased understanding of the actual network by comparing with its multi-layer functional network counterpart.

In the existing studies on PTNs using graph theory, the networks are represented as regular graphs with nodes representing the bus stops and edges representing the routes connecting the bus stops. Depending on the edge type, the graph can be modeled in either L-Space, where an edge exists between two nodes or stops if they are consecutive stops in a route or P-Space, where an edge exists between all node pairs that are serviced by a specific route [4]. Although this approach of PTN analysis using regular graph representation is simple, it has been extremely successful. However, with the advancement of research in the complex network analysis, a representation beyond these simple graphs (e.g., directed, weighted, bipartite/multipartite etc.) is needed to investigate the complicated and realistic network behavior. To contribute to one such representation, in this paper, we propose a novel approach called supernode graph structure representation to analyze the PTN structure of the London city. Unlike the regular graph representation, in supernode graph representation we combine geographically closely associated nodes based on a specific criterion, resulting in a more compact representation which benefits in a more realistic network analysis. To model the PTN graph in London, we have considered the bus-stops, overground stops, underground stops and the DLR (Docklands Light Railway) stops. For the multi-layer PTN analysis, the bus stops in the city are treated as one transport layer and is termed as Bus Transport Network (BTN) layer, the underground, the overground and the DLR stops are treated as the other transport layer termed as Metro transport (MTR) layer. Initially, the London PTN structure is represented as a regular graph structure (in L-Space) which is then modified to supernode graph structure at individual layers. The two individual layers are integrated by the proposed method of spatial amalgamation. A node weight approach is used to assign weights to the nodes which helps identify highly influential nodes in the network with respect to both mono-layer and the integrated multi-layer networks.

II. Supernode Graph Structure Representation

A graph G is a set of nodes V and edges E , i.e., $G = (V, E)$. Considering the spatial analysis of the network, in the current work, a graph G is represented by $G = (V(x, y), E)$ where $V = \{n_i(x_i, y_i), x_i = \text{latitude}, y_i = \text{longitude}\}$ and $E = \{e_{ij} \rightarrow (n_i(x_i, y_i), n_j(x_j, y_j))\}$. $N = |V|$, indicates

978-1-5386-4881-0/18/$31.00 ©2018 IEEE



(a) (b) (c)

Fig. 1. (a) Spatial location of bus stops and metro stops in the London public transport network, (b) *Buffer* zone of 500 m around a node in MTR layer along with the overlapped nodes on BTN layer, (c) BTN, MTR and the coupling/intermediate layer in the analysis of the London PTN.

the network size. In the subsequent sections, for simplicity, $n_i(x_i,y_i)$ is represented as n_i assuming that a particular node is always identified with its latitude and longitude. An $N \times N$ adjacency matrix A with entries a_{ij} is used to describe the connectivity between nodes, where a_{ij}=1 if there exists a route between nodes n_i and n_j, and 0 otherwise. The geographical locations of the stations are represented in WGS84 datum standard using ArcGIS tool [5]. Fig. 1(a) shows the spatial locations of public transport stops in the London city with 20553 nodes and 29561 edges.

In transport network analysis, the inspection of spatial embedding of nodes has resulted in a new type of network element called *supernode*. A supernode is a set of geographically closely associated nodes which satisfy the condition, $d_{ij} < d_{th}$, where d_{ij} is the geographic distance between two nodes n_i and n_j, and d_{th} is a threshold distance. The value of d_{th} is set to be 100 m in this paper assuming that it is a walkable catchment to reach a station. The distance d_{ij} is evaluated using the Haversine formula [6]. The set of nodes satisfying the condition $d_{ij} < d_{th}$ are combined to represent a single node called the supernode. The combining of nodes is not physical, but is a structural re-organization, which yields in a compact topological representation and aids in a more practical network analysis. A Supernode graph structure consists of regular nodes (V), supernodes (V_S), regular edges (E) and superedges (E_S), i.e., $G = (V(x,y), V_S(x,y), E, E_S)$, where V_S and E_S are given by (1) and (2) respectively.

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

where $N_S = |V_S|$, $sn_i = \{n_j, n_k\}$, such that $(d_{jk} < d_{th})$, i.e., The supernode set (V_S) is a set of supernodes where each supernode is a combination of two or more regular nodes with a geographic distance less than d_{th}, with $d_{th} = 100$ m. If $(sn_i \cap sn_j) \neq \emptyset$, then, $\tilde{sn_i}$=$\{sn_i \cup sn_j\}$, i.e., there exists a condition where some nodes are common to multiple supernodes and such nodes aid in combining the supernodes together to form a giant supernode ($\tilde{sn_i}$), which is assigned with a unique node ID. The newly formed supernode is assigned with a new spatial location, which is the mean location value of the corresponding supernode elements, i.e., $sn_i(x_i, y_i) = \{n_j(x_j, y_j), n_k(x_k, y_k)\}$, where $x_i = \text{mean}(x_j, x_k), y_i = \text{mean}(y_j, y_k)$. The superedge set is defined as

$$E_S = \{e_{sn_i,sn_j} \cup e_{n_i,sn_j}\} \; \forall sn_i, sn_j \in V_S, n_i \in V \tag{2}$$

Assuming that an edge e_{ij} exists between nodes n_i and n_j, a superedge can then be defined between two supernodes as $e_{sn_i,sn_j} \rightarrow (sn_i, sn_j)\forall(n_i \in sn_i, n_j \in sn_j)$, or a superedge can be defined between a regular node and supernode as $e_{n_i,sn_j} \rightarrow (n_i, sn_j)\forall(n_i \in V, n_j \in sn_j)$. While defining the supernode structure, some of the original nodes and self loops will be eliminated due to the formation of supernodes and superedges. Thus, using supernode representation, a network can be structurally defined close to its original network with a reduced data set.

III. Multi-layer Network

Considering the supernode graph structure generated in Section II for the BTN and MTR layers, we define the multilayer network, M, as follows [2]

$$M = \left(V_M, E_M, \tilde{V}, \mathrm{L}\right) \tag{3}$$

where $\tilde{V}$ is the node set containing both regular nodes and supernodes including all the layers. $\mathrm{L} = \{L_a\}_{a=1}^{d}$ is the set of elementary layers defined by d aspects 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 this paper, $d = 2$, with an elementary layer and an additional layer. Also, $V_M \subseteq \tilde{V}$ such that $V_M \times L_1, \times L_2 \times .. \times L_d$ is a node set in a the multi-layer network M with different layers L_1 to L_d. $E_M \subseteq V_M \times V_M$ is the edge set containing both regular edges and superedges including all the layers. Among the BTN and MTR layers, the BTN layer is considered the elementary layer since the London statistics [7] indicate that 54$\%$ of the population prefers bus transport mode for their daily commute and MTR layer is considered as the additional layer or the supporting layer with 27$\%$ population using metro services. The remaining 18$\%$, who prefers national rail services is not considered in the analysis since the combined bus and metro services have covered up to 80$\%$ of the daily commuting modes. It is interesting to observe that, the transport network structures

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

belong to the category of layer-disjoint multi-layer networks where each node exists in at most one layer [1].

$$(n_i)_\alpha, (n_i)_\beta \in V_M \Rightarrow \alpha = \beta \tag{4}$$

where a node is present either in layer α (i.e., $L_1 = \alpha$) or β (i.e., $L_2 = \beta$). The layer-disjoint property signify an important observation that there exists no edges between the layers in the actual network structure, and the layers are normally connected virtually (by a small walking distance) and not physically. Hence, to integrate the two layers, in this paper, we employ the method of spatial combining of nodes. The spatial combining is carried out using the *buffer* feature in ArcGIS tool where a geographical area with a radius of 500 m is considered with a node in the MTR layer as a central point, and the nodes in the BTN layer that overlaps with the region considered (500 m radius) are extracted as shown in Fig 1(b). According to the London statistics [8], the walkable catchment for bus/tram and MTR stations are 400 m and 700 m respectively. In this paper, we consider 500 m as a walkable distance for interchanging between different transport layers. The overlapped nodes in the 500 m region around a node in the MTR layer are the stations that allow passengers to interchange between the layers and these nodes are treated as the third layer called the intermediate layer as shown in Fig 1(c). The set of nodes in the intermediate layer is a subset of the BTN layer which aids in better understanding of the virtual connectivity between the BTN and MTR layers. The intra-layer edge sets E_α and E_β, and the inter-layer or coupling edge set E_C is defined as

$$E_\alpha = \{e_{ij}\} \;|\; e_{ij} \rightarrow (n_i, n_j)_\alpha \;\; \forall n_i, n_j \in V_\alpha \tag{5}$$

$$E_\beta = \{e_{kl}\} \;|\; e_{kl} \rightarrow (n_k, n_l)_\beta \;\; \forall n_k, n_l \in V_\beta \tag{6}$$

$$E_C = \{e_{ik}\} \;|\; e_{ik} \leftrightarrow (n_i, n_k), n_i \in V_{\text{zone}}, n_k \in V_\beta \tag{7}$$

where α and β are the BTN and MTR layers, respectively; E_α and E_β are the edge set of layer α and β, respectively; V_β is the node set of layer β; and V_{zone} is the set of overlapped nodes in the 500 m zone such that $V_{\text{zone}} \subseteq V_\alpha$.

IV. Node Weight Analysis in Multi-Layer Network

The accessibility of a bus stop by the public is greatly influenced by the presence of points of interests (POIs) around the public transport station, proximity of the transport station to a POI (Public Transport Accessibility Level) and its connectivity (degree). Information about location of POIs around a station, the knowledge about demographics and job opportunities contribute to a better estimation of the demand serviced by a node in the PTN analysis. In order to analyze the demand distribution across multi-layers, in this section, we propose an approach to assign node weight to a node considering mono-layer and multi-layer analysis. To make the analysis more reasonable, we consider the smallest geographical division. For example, the land area in the London city is divided into $Boroughs$, which are subdivided into *wards*. By understanding the distribution patterns of POIs, working population and job opportunities in every ward, we estimate the demand distribution at microscopic level which aids in the identification of highly influential nodes in the network (according to their usage) with respect to both mono-layer and multi-layer analysis. In the following section we discuss the detailed procedure for assigning a node weight.

Step 1: Extract information regarding land area, working population, available job opportunities and the list of POIs in each ward. The detailed information of the 624 wards in London can be extracted from [9].

Step 2: By using the *spatial combining* feature in ArcGIS, obtain the count of POIs, bus stops and metro stops in each ward, and calculate their densities (e.g., number of POIs per km^2 in a ward)

Step 3: Evaluate the node occupying probability (NOP), which is defined as the number of people using a particular station in each ward. For a reasonable analysis, we calculate NOP as

$$(\text{NOP}_{i,\alpha})_Z = (\rho_{P_\alpha}/\rho_{N_\alpha})_Z \tag{8}$$

where, $\text{NOP}_{\text{i},\alpha}$ is the node occupying probability of station i on layer α, ρ_{P_α} is the density of working population accessing the stations on layer α, and ρ_{N_α} is the density of public transport stations on layer α in ward Z, for Z = 1, 2, .., 624. Extending (8) to layer β, $(\text{NOP}_{\text{i},\beta})_Z = \left(\rho_{P_\beta}/\rho_{N_\beta}\right)_Z$. As per the statistical data from the London government [10], the number of people accessing a particular station in a ward on a particular layer is derived from the total working population count as

$$(P_\alpha = 0.54 * P_T)_Z, \quad (P_\beta = 0.27 * P_T)_Z \tag{9}$$

where P_α and P_β denote the working population accessing the stations on layer α and β respectively, and P_T denotes the total working population in a particular ward.

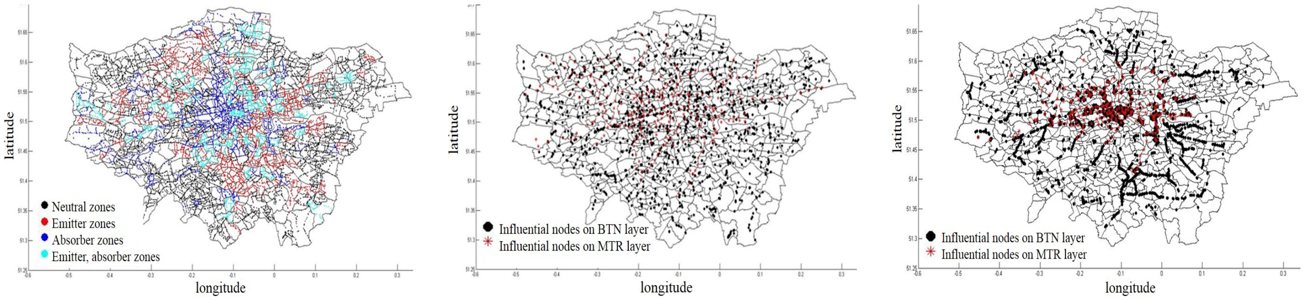
Step 4: Categorize each ward as either an emitter, absorber or neutral region based on the three information categories, i.e., POIs, working population count and the job opportunities in a ward as shown in Table I. From the distribution patterns of the three categories, the distribution of POIs and the job opportunity in a given ward follow an exponential distribution, and the distribution pattern of working population follows a normal distribution. In Table I, the category type in a particular ward is considered scarce if its value is less than the mean value of the normal distribution for working population category or the median value of the exponential distribution for POIs and job opportunity categories. Otherwise, it is considered abundant.

Step 5: Assign a node weight to each station

$$(w_{i_\alpha})_Z = \left(\frac{\rho_{P_\alpha}}{\rho_{N_\alpha}}\right)_Z + k_{i_\alpha} \forall Z = 1, 2, .., 624 \tag{10}$$

where $w_{i_\alpha Z}$ is the weight of node i on layer α in a particular ward, $(\frac{\rho_{P_\alpha}}{\rho_{N_\alpha}})_Z$ is the NOP of a station in a given ward Z, and k_{i_α} is the node degree which indicates the connectivity of a node on layer α when the network is analyzed as a mono-layer network. Equation (10) holds for the β layer network as well.

Step 6: Normalize the node weight to ensure the data integrity



(a) (b) (c)

Fig. 2. (a) The wards in London city categorized as emitter, absorber or neutral zones along with the spatial location of bus stops, (b) influential nodes in the London PTN under mono-layer analysis, (c) influential nodes in the London PTN under multi-layer analysis.

in all wards. The closer the value of normalized node weight (w_{i_norm}) to one, higher is the usage of a node. Considering the mono-layer analysis, the most influential nodes in the network at different layers are the nodes with their normalized node weight equal to 1. Fig. 2(a) shows the spatial locations of the stations on layer α along with their ward type categorized as emitter or absorber or neutral type. Fig. 2(b) shows the influential nodes according to the node weight assigned for mono-layer network analysis. Since we normalize the node weight in every ward, there exists at least one influential node in each ward. Here influential nodes refers to the most used nodes in the network at the microscopic level (wards). Thus, assigning weights to the nodes not only aids in identifying the existing influential nodes, but also assists in predicting the future influential nodes in the network with the increase in the number of POIs, job opportunities, and population distribution around a station in a chosen ward. However,

TABLE I
CLASSIFICATION OF WARDS INTO EMITTER/ABSORBER/NEUTRAL REGIONS.

Job opportunities	Working population	POIs	Ward 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

the node weight assigned using (10) may slightly bias our results since people might choose multiple transfer modes for commuting instead of using a mono-layer mode. To take this into consideration, the independent mono-layer network is collapsed into a dependent multi-layer network by integrating both layers as discussed in Section III, and thus (10) is modified as

$$w_i = w_{i_\alpha} + C_{b_i} \tag{11}$$

where w_i is the overall weight of a node i, w_{i_α} is the weight of a node i on layer α ((11) holds good for layer β also), and C_{b_i} is the betweenness centrality of node i considering the integrated network and is given by $C_b(i) = \sum\limits_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$, where σ_{jk} is the total number of shortest paths between nodes j and k, and $\sigma_{jk}(i)$ is the shortest paths between nodes j and k passing through node i. Fig. 2(c) shows the influential nodes in the network according to the node weight assigned with integrated network analysis. As observed from Fig 2(b), determining the set of influential nodes in a network by analyzing its node weight according to mono-layer analysis is different from that using multi-layer analysis as shown in and 2(c). Hence, the consideration of the inter-connectedness between the transport layers assist in understanding the dynamic network behavior.

V. CONCLUSION

As compared to regular graph representation, the proposed supernode graph representation offers a more practical way of analyzing the transport network topology which not only aids in a realistic network analysis, but also accounts for a precise network representation without significant loss. The method of spatial integration using the concept of walkable catchment is employed to integrate the BTN and MTR layers in the London city. A node weight analysis method is proposed to assign weight to a node. It is observed that, the assigned node weight differ significantly under the mono-layer and multi-layer analyses of the network which indicates that neglecting the interaction between the transport layers may bias our understanding of the overall network behavior considering the real-world usage of the network. The classification of microscopic regions in a city into different types such as emitter, absorber or neutral regions according to POIs, working population, and the job opportunities further aids in understanding the behavior of an influential node in the network.

VI. ACKNOWLEDGMENT

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

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

REFERENCES

[1] 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.

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

[3] 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.

[4] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, “Public transport networks: empirical analysis and modeling,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 68, no. 2, pp. 261–275, 2009.

[5] M. H. Price, *Mastering ArcGIS*. McGraw Hill Higher Education, 2004.

[6] C. Veness, “Calculate distance and bearing between two latitude/longitude points using haversine formula in javascript,” *Movable Type Scripts*, 2011.

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

[8] S. Cooper, “Measuring public transport accessibility levels (ptals),” Technical Report, Transport for London, pp. 1–7, 2003.

[9] “Borough Profiles London Dataset,” https://data.london.gov.uk/dataset/london-borough-profiles/, [Online; accessed 10-Jan-2017].

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

978-1-5386-4881-0/18/$31.00 ©2018 IEEE