Physica A 389 (2010) 5852–5863

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Weighted complex network analysis of travel routes on the Singapore public transportation system

Harold Soh [a,*], Sonja Lim [a], Tianyou Zhang [a], Xiuju Fu [a], Gary Kee Khoon Lee [a], Terence Gih Guang Hung [a], Pan Di [c], Silvester Prakasam [c], Limsoon Wong [b]

[a] *Advanced Computing, Institute of High Performance Computing, Singapore*

[b] *National University of Singapore, Singapore*

[c] *Land Transport Authority Singapore, Singapore*

ARTICLE INFO

Article history:
Received 25 May 2010
Received in revised form 9 August 2010
Available online 19 August 2010

Keywords:
Weighted complex network analysis
Topological analysis
Transportation system

ABSTRACT

The structure and properties of public transportation networks have great implications for urban planning, public policies and infectious disease control. We contribute a complex weighted network analysis of travel routes on the Singapore rail and bus transportation systems. We study the two networks using both topological and dynamical analyses. Our results provide additional evidence that a dynamical study adds to the information gained by traditional topological analysis, providing a richer view of complex weighted networks. For example, while initial topological measures showed that the rail network is almost fully connected, dynamical measures highlighted hub nodes that experience disproportionately large traffic. The dynamical assortativity of the bus networks also differed from its topological counterpart. In addition, inspection of the weighted eigenvector centralities highlighted a significant difference in traffic flows for both networks during weekdays and weekends, suggesting the importance of adding a temporal perspective missing from many previous studies.

1. Introduction

Because the structure of a network often affects its function [1], deciphering the topology and dynamics of the underlying networks is a prerequisite to a full understanding of connected, interacting systems. For the past half century, research on networked structures has focussed primarily on Erdös and Rényi (ER) random graphs, the canonical description of complex networks. However, ER random networks are theoretical constructs and may only represent a small subset of real-world systems.

The problem of real-world network analysis is that of complexity. Interesting real-world networks can consist of millions of nodes connected by a complicated set of edges, making them difficult to unravel. Fortunately, advances in complex network theory, measurement techniques and computational power have greatly improved our ability to analyze such structures. In the past few years, we have made fascinating discoveries on the nature of a diverse set of complex systems from the neural network of the nematode *Elegans* [2] to the World Wide Web (WWW) [3].

In this paper, we contribute a *complex network analysis of passenger travel routes* on the rail and bus public transport systems in the island nation of Singapore. The bus network studied in this work is among the largest studied to date, with more than 4130 nodes. Moreover, the Singapore Land Transport Authority uses an integrated distance-based fare

* Corresponding author. Tel.: +44 (0)20 7594 6323.
E-mail address: haroldsoh@imperial.ac.uk (H. Soh).

0378-4371/$ – see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.08.015

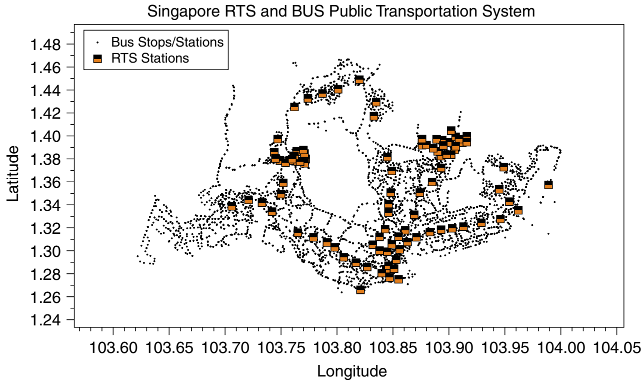


Fig. 1. The Singapore public transportation system, comprising the RTS and BUS systems.

system, which captures data *both* when commuters *board* and *alight* busses through the use of smart card readers. Alighting information on bus networks is usually unavailable or based on estimation in other urban transport information systems (e.g. London, Paris and New York). As a result, the public transport data captured through the Singapore public transport ticketing system is one of the most comprehensive in the world.

We compare its characteristics to previously studied transportation systems such as the Boston subway [4], the Indian railway [5], Chinese railway [6], the world-wide airport network [7] and the public transportation system in [8]. In addition to a topological study, we also investigate the dynamical properties of the networks by incorporating the magnitude of interactions [9–14]. We find that the Singaporean transport systems share several similar (but not identical) dynamical features with the world-wide air transportation network [10] and the German and Indian railway networks [15]. Furthermore, we extend previous analysis into the temporal domain by considering the differences in properties over weekdays and the weekend.

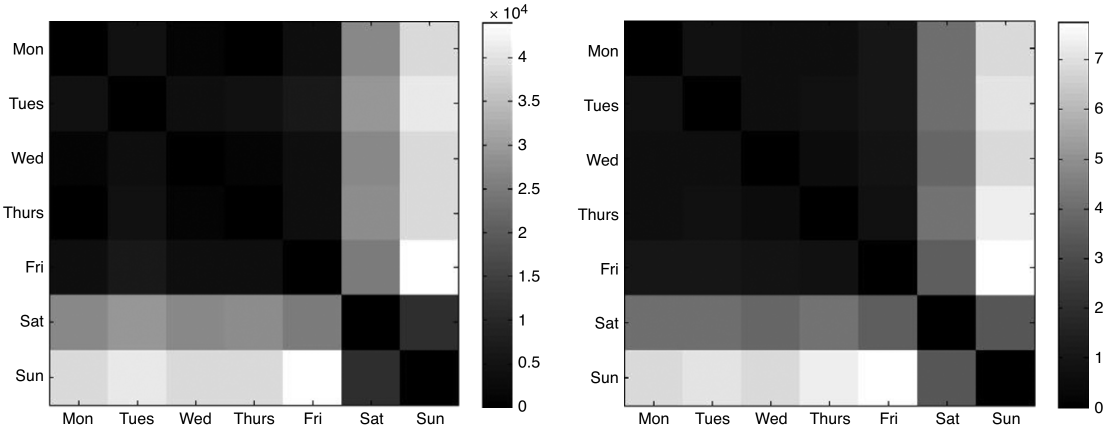
In the next section, we will set the stage by describing the Singapore Public Transportation system and the dataset provided by the Land Transport Authority. In Section 3, we report on the degree, strength, clustering, assortativity and eigenvector centrality characteristics of the transportation networks. In brief, our study highlights the importance of studying both the topological and dynamical properties of networks. For example, from a topological perspective, the Singapore rail network appears similar to a highly connected ER random graph (indicating travel between almost all districts in Singapore). However, a dynamical analysis reveals a more complex scale-free system in which certain regions enjoy substantially greater traffic. Furthermore, an eigenvector analysis (in Section 3.5) shows that the traffic can differ significantly depending on the day of the week, suggesting the importance of temporal effects. We believe that our results will have an impact on future modelling and simulation studies, for example in epidemiology, where static (rather than dynamic) networks are the norm. We elaborate upon these aspects as well as avenues for future work in Section 4, which concludes our paper.

2. Case study: the public transportation system in Singapore

Singapore is an island nation at the southern tip of the Malay Peninsula in South East Asia. Home to approximately 4.86 million people but with a land area of only 710.2 km^2, Singapore is the third densest country in the world [16]. Singapore has experienced tremendous economic growth since its independence in 1965; the International Monetary Fund now ranks Singapore as the fifth wealthiest country in the world in terms of Gross Domestic Product per capita [17]. Part of its rapid economic progress can be attributed to an efficient transportation system.

Pre-World War II, the human-powered trishaw was the main means of public transportation in Singapore. Today, the land-based public transportation system in Singapore is comprised of two efficient, sophisticated networks: (1) the rail or Rapid Transit System (RTS) and (2) the bus system (BUS), shown in Fig. 1. The RTS network in 2008 consisted of 93 stations (grouped into three mass-rapid transit (MRT) lines and three light-rail (LRT) lines), connecting all major districts across the island. The public BUS network is larger, with more than 4000 bus stops covering almost all populated regions. In 2007, approximately 4.5 million trips were made on the RTS and BUS systems daily.

The main payment method employed by the Singapore public transportation system is a contact-less smart card system called EZ-Link [18]. An individual taps his or her smart card once upon entry to the RTS station (or BUS) and once more upon exit (or alighting). In the case of the RTS, individuals can change trains at intermediate hub stations without exiting the RTS network. It is through the EZ-Link payment system that we are able to capture the traffic serviced by these transportation networks. Our datasets that are provided by the Land Transport Authority of Singapore (LTA) list the daily in and out traffic for each RTS station and bus stop to all other stations and stops. The data spans one week (Monday to Sunday) in January 2008, capturing 10.6 million RTS passenger rides and 19 million BUS passenger rides.



(a) RTS daily weight difference.

(b) BUS daily weight difference.

Fig. 2. Daily edge weight difference across days in a week for the Singapore Rapid Transit System (RTS) and Singapore bus (BUS) networks. Each cell represents the mean squared difference (m.s.d.) in weights between two days in the week studied. Lighter cells represent a greater difference in m.s.d. We can clearly see two clusters: one comprising the weekdays (Monday to Friday) and the other comprising the weekends (Saturday and Sunday).

2.1. Data preprocessing and graph generation

We transformed the list format data obtained from the LTA into seven daily graphs. Travel for each day was represented as a weighted graph G with N nodes and M edges, an associated adjacency matrix $\mathbf{A} = [a_{ij}]$ and a weight matrix $\mathbf{W} = [w_{ij}]$ representing the number of passengers travelling between locations i and j in a single day.

Due to the method of data capture, it is important to note that we are not studying the underlying physical structure of the networks but of the movement of people between the different nodes. As such, when we say two nodes i and j are connected, $a_{ij} = 1$, we mean that there is at least one passenger travelling between these two nodes during the week. Also, the bus and rail networks are disconnected, since transfers from the RTS to the BUS network (and vice versa) are not present in the datasets.

There was a significant difference between edge weights during weekdays and weekends on both the BUS and RTS networks (see Fig. 2). To simplify our analysis, we combined the Monday to Friday graphs for the rail and bus networks into weekday graphs by averaging the weights along each edge. We created the weekend networks in the same manner (averaging the Saturday and Sunday graphs). In addition, we made the assumption that typical travel was bi-directional, and hence each graph was made undirected ($w_{ij} = w_{ji}$) by averaging the in and out edge weights. Finally, we were left with four networks: RTS_D and RTS_E representing passenger travel on the RTS system during the weekdays and weekend, respectively, and, likewise, BUS_D and BUS_E for the BUS system.

3. Complex weighted network analysis of travel on the Singapore public transportation networks

In this section, we present a topological and dynamical analysis of the RTS_D, RTS_E, BUS_D and BUS_E networks. Table 1 shows all computed network statistics, from basic network properties such as the number of nodes and edges to the more complex metrics such as clustering, assortativity and eigenvector centrality.

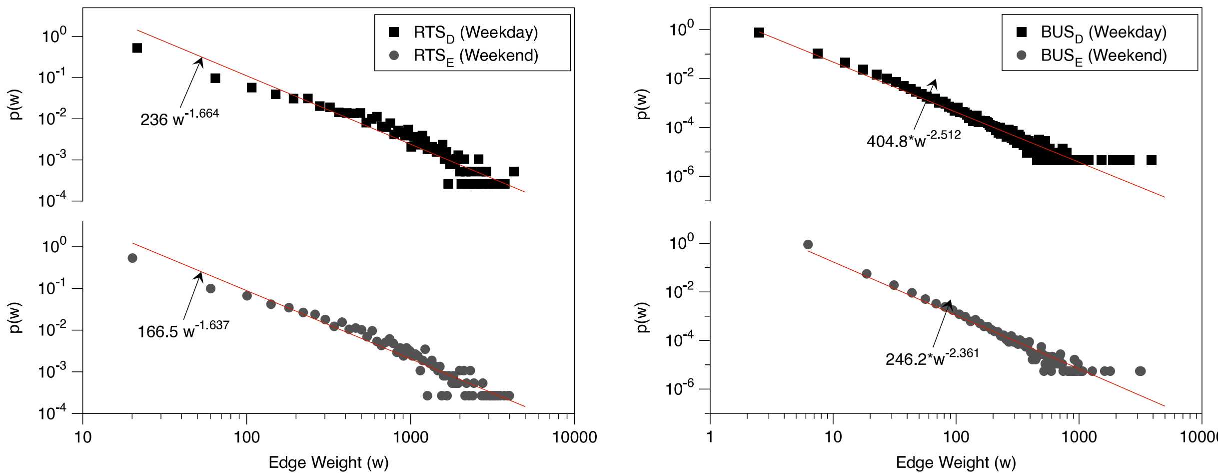
3.1. Basic properties and path length

The final RTS weekend and weekday networks both have 93 nodes and a similar number of edges, $M = 3843$ (RTS_D) and $M = 3733$ (RTS_E). As expected, BUS_D and BUS_E are significantly larger, with 4134 and 4142 nodes respectively. Although the weekend BUS network possessed more nodes, the weekday BUS network had almost 51×10^4 more edges ($M = 213{,}103$ (BUS_D), compared to BUS_E where $M = 180{,}109$), suggesting less travel during the weekends but to more distinct locations.

All four networks feature similar small average path lengths (path length being the minimum number of edges traversed to get from one node to another), $\langle l \rangle \approx 1.1$ (RTS) and $\langle l \rangle \approx 2.5$ respectively (BUS). These path lengths are similar to those found for the transportation routes for cities in Poland [8]. The RTS and BUS networks also feature small diameters (maximum path length of a network) $d = 2$ (both RTS_D and RTS_E) and $d = 4$ (BUS_D) and $d = 5$ (BUS_E). For the RTS network, a small average path length of a single hop means that there is travel between almost all rail stations of Singapore, regardless of geographical distance. Moreover, we observed that the distance travelled using buses was mostly short, with 95% of all rides being below 10 km (as shown in the cumulative probability distribution of the distances travelled between pairs of nodes plotted in Fig. 4). In contrast, more than 50% of all rides on the RTS system were above 10 km.

Table 1
Statistical properties of the RTS and BUS weekday and weekend networks.

Property	RTS_D	RTS_E	BUS_D	BUS_E
Number of nodes, N	93	93	4131	4139
Number of edges, M	3843	3733	213103	180109
Average shortest path, $\langle l \rangle$	1.101	1.127	2.5403	2.5762
Diameter, d	2	2	4	5
Average weight, $\langle w \rangle$	206.354	176.456	6.434	6.462
Weight range	(0.1, 4292.9)	(0.25, 4019.25)	(0.1, 2583)	(0.25, 3124)
Average degree, $\langle k \rangle$	82.6452	80.2796	103.172	87.03
Degree range	(35,92)	(29, 92)	(1, 1073)	(1, 1048)
Average strength, $\langle s \rangle$	17054.2	14165.8	663.803	562.393
Strength range	(18.20, 65451.3)	(84.25, 59899)	(0.2, 64236.1)	(0.25, 56121)
Assortativity, r	−0.0875	−0.0775	0.054997	0.0146622
Average clustering, C	0.9341	0.9216	0.562047	0.533689
Average weighted clustering, C^w	0.9785	0.9716	0.655622	0.636493
Average centrality, X	0.103	0.103	0.0104	0.0102
Average weighted centrality, X^w	0.0745	0.0735	0.00284	0.00303



(a) RTS weight distribution. (b) BUS weight distribution.

Fig. 3. Distribution of passenger travel (weights) on the RTS and BUS weekday and weekend networks, along with power-law fits.

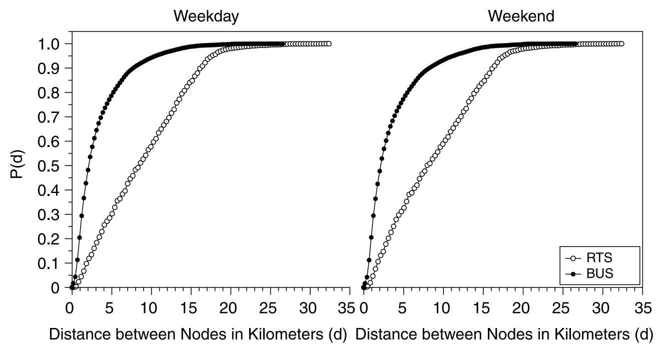
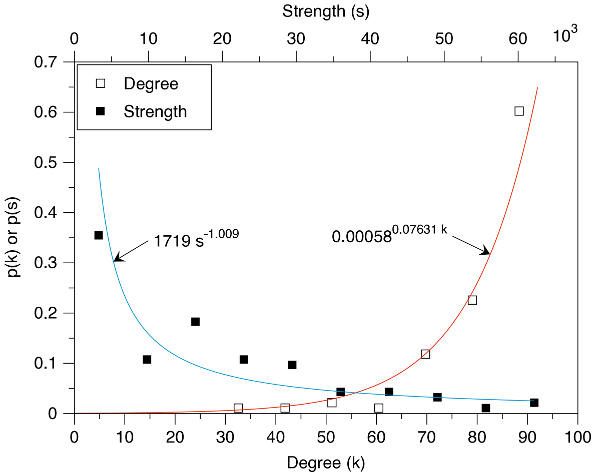
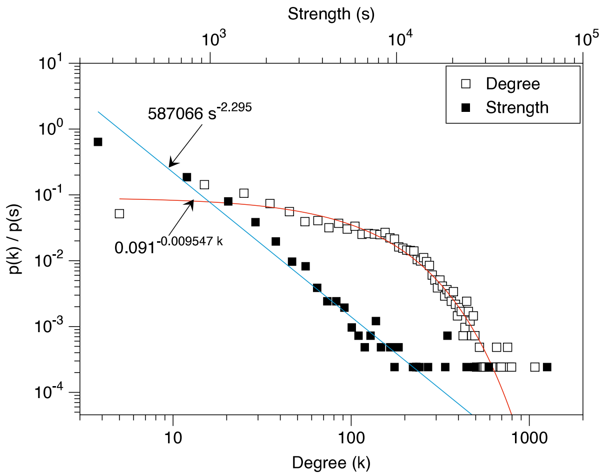


Fig. 4. Cumulative distribution of distances travelled on both the RTS and BUS networks. We observed the distances travelled on buses to be shorter.

Turning our attention to the number of rides (weight) along each edge, we observed that the traffic between rail stations varied greatly: $w \in (0.1, 4293)$ for RTS_D and $w \in (0.25, 4020)$ for RTS_E. Similarly large ranges are present in the BUS networks, where $w \in (0.1, 2583)$ for BUS_D and $w \in (0.25, 3124)$ for BUS_E. The weight distributions for the RTS and BUS networks shown in Fig. 3(a) and (b) appear to follow a power-law distribution $p(w) \sim w^{-\gamma}$, where $\gamma \approx 1.6$ for RTS and $\gamma \approx 2.5$ for BUS, indicating the presence of travel routes with very high traffic, $w_{ij} > 2000$ rides per day. The natural question to follow is whether weights are distributed independently across the edges or whether some nodes enjoy "prominence" in the network, which we address in the next section.



(a) RTS_D (weekday).



(b) BUS_D (weekday).

Fig. 5. Degree and strength distributions for the RTS and BUS weekday networks.

3.2. Degree and strength distribution

The degree of a node is a measure of its connectivity, and its strength can be viewed as an indication of its centrality [14, 10]. For a given node i, its *degree*, $k_i = \sum_j^N a_{ij}$, is the number of nodes it is linked to. We can average this over all nodes to give the graph's average degree:

$$\langle k \rangle = \frac{1}{N} \sum_i^N k_i = \frac{1}{N} \sum_i^N \sum_j^N a_{ij}. \tag{1}$$

Subsequently, a node's *strength* is simply the sum of the weights on the edges incident upon it, $s_i = \sum_j^N a_{ij} w_{ij}$, and the graph's average strength is given by

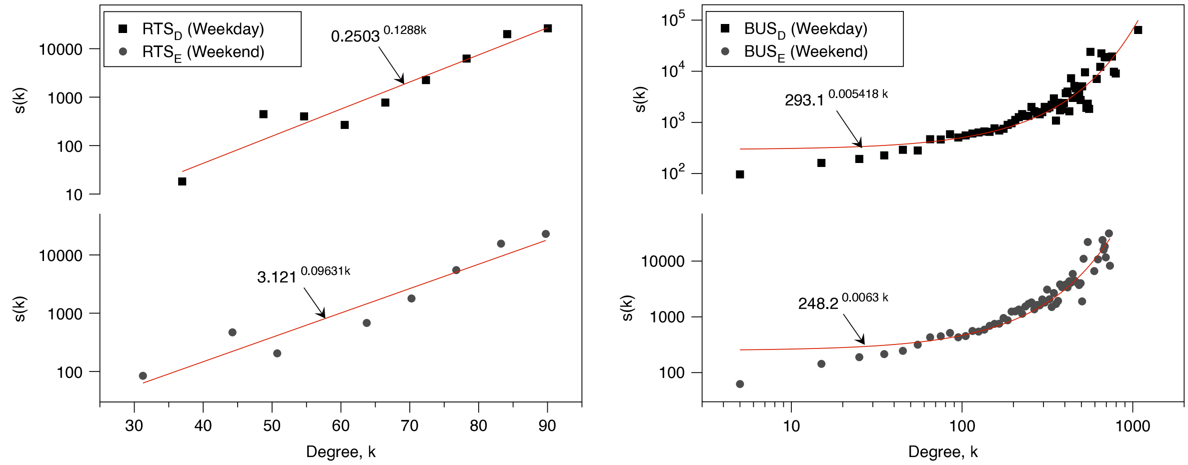
$$\langle s \rangle = \frac{1}{N} \sum_i^N s_i = \frac{1}{N} \sum_i^N \sum_j^N a_{ij} w_{ij}. \tag{2}$$

Both the RTS weekday and weekend networks possess high average degrees, $\langle k \rangle = 82.65$ (RTS_D) and $\langle k \rangle = 80.28$, ($RTS_E$), indicating high connectivity among the rail stations. The degree distribution for RTS_D (see Fig. 5(a) showed that $p(k)$ increases exponentially with k, and a majority of nodes (80%) possess high degree $k_i > 80$, indicating that a majority of RTS nodes exist in a highly connected cluster (RTS_E has a similar distribution). That said, the strength distributions (filled markers) revealed that although many nodes share similarly high degree, the traffic handled by each rail station differed significantly. In fact, the strengths of stations appear scale-free (indicating the existence of hub nodes with very high traffic) and follow a power-law distribution $p(s) \sim s^{-\gamma}$ with $\gamma \approx 1$ for both RTS_D and RTS_E.

The BUS system has a topology different from that of the RTS network. Relative to the size of the network, $N \approx 4100$, the degrees of the bus stops are small, with $\langle k \rangle = 103$ (BUS_D) and $\langle k \rangle = 87$ (BUS_E). The weekday degree distributions for BUS_D (Fig. 5(b)) appear exponentially distributed, $p(k) \sim \alpha^{\beta k}$, where $\beta \approx -0.01$. The exponential distribution of the degrees suggest that network's connectivity evolved randomly, possibly due to the many factors that affect urban development, as suggested by Sienkiewicz and Holyst [8].

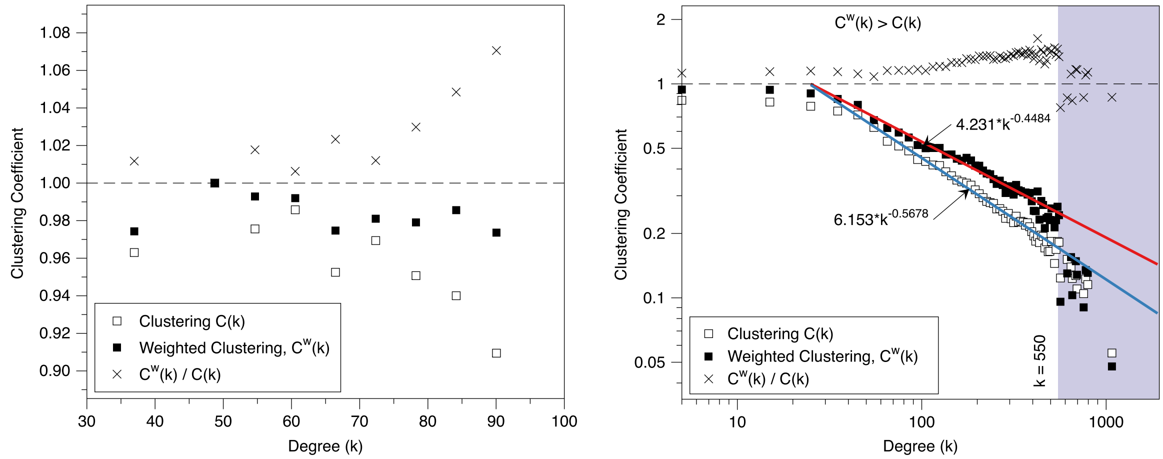
On the other hand, the strengths of the BUS nodes follow a power-law distribution (at least for $s < 20{,}000$), with exponents $\gamma \approx 2$. Similar to the RTS networks, there exist high-traffic hub BUS nodes. Therefore, although connections between pairs of nodes may be random, the magnitudes (or concentrations) of the travel routes are not. Instead, if we take the perspective that weights are multi-edges [12], then the edges (or travel routes) appear to emerge out of a preferential attachment process. In other words, an individual is more likely to travel to a station that other people are travelling to.

The strength spectrums (obtained by averaging the strength $s(i)$ over all nodes with a given degree) for both the RTS and BUS networks (Fig. 6) illustrate a positive relationship between the degree and the strength of nodes. In fact, the strengths grow exponentially with degree for both the rail and bus transportation systems. This observation fits with the intuition that the more connected a station is, the more traffic it handles, but also implies that the traffic grows much faster than the number of connections (Fig. 6).



(a) RTS. (b) BUS.

Fig. 6. Strength (as a function of degree) of the RTS and BUS networks.



(a) RTS$_D$ (weekday). (b) BUS$_D$ (weekday).

Fig. 7. Clustering spectrum for the RTS and BUS networks.

3.3. Topological and dynamical clustering

The clustering coefficient is a measure of cohesiveness around a given node i, and it is defined by the equation

$$C_i = \frac{2E_i}{k_i(k_i-1)} = \frac{2}{k_i(k_i-1)} \sum_{j,h} a_{i,j}a_{i,h}a_{j,h}, \tag{3}$$

where E_i is the number of edges between node i's neighbors and $k_i(k_i-1)/2$ is a normalization factor equal to the maximum number of possible edges among the neighbors. Because of this normalization, C_i is in the interval [0, 1], where 0 and 1 indicate that none or all of node i's neighbors are linked, respectively. Averaging C_i over all the nodes in the network gives the average clustering coefficient $\langle C\rangle = \frac{\sum_i^N C_i}{N}$, a convenient summary statistic for cohesiveness.

The range of clustering coefficients C_i for both RTS$_D$ and RTS$_E$ are similarly high and narrow, $C_i \in (0.90, 1)$ (RTS$_D$) and $C_i \in (0.87, 1)$ (RTS$_E$), indicating that the network is highly clustered, as expected, given the degree distributions previously discussed. The average clustering coefficients $\langle C\rangle$ for the weekend and weekday RTS networks are 0.934 and 0.921, respectively, slightly higher than the value from an equivalent ER random graph, $\langle C\rangle_{ER} = 0.857$, given by

$$\langle C\rangle_{ER} = \frac{(\langle k^2\rangle - \langle k\rangle)^2}{N\langle k\rangle^3}. \tag{4}$$

Compared to previous studies, the RTS networks are significantly more cohesive than the Indian railway ($\langle C\rangle_{IR} = 0.69$) [5], the public transportation networks in Poland ($0.68 < \langle C\rangle_P < 0.85$) [8], the location network of Portland ($\langle C\rangle_{PL} = 0.0584$) [19] and the Chinese railway ($\langle C\rangle_{CR} = 0.835$) [6].

From Table 1, we can see that the average clustering coefficients for the BUS networks are smaller than for the RTS networks with $\langle C = 0.56\rangle$ (BUS_D) and $\langle C = 0.53\rangle$ (BUS_D). That said, the clustering values are approximately 30 times larger than for an ER random graph of the same size (0.017 for weekdays and 0.013 for the weekend). Together with their small average path lengths (see Section 3.1), this high clustering coefficient indicates that the BUS networks are small-world.

While C_i measures only topological cohesiveness, the *weighted* clustering coefficient C_i^w takes into account the weights of edges,

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{i,j} + w_{i,h}}{2} a_{i,j} a_{i,h} a_{j,h}, \tag{5}$$

with the average weighted clustering coefficient given by $\langle C^w\rangle = \frac{\sum_i^N C_i^w}{N}$. In the case where the weights are completely uncorrelated, $C^w = C$ and $C^w(k) = C(k)$ [10]. However, weights in real-world networks are often correlated, leading to two possible situations.

1. $\langle C^w\rangle > \langle C\rangle$: in this case, closed triangles are more likely formed by edges with larger weights, i.e., clustering is formed by edges with larger weights.
2. $\langle C^w\rangle < \langle C\rangle$: closed triangles are more likely formed by edges with smaller weights and, as such, clustering is formed by edges with low weights.

In Barrat's study of weighted networks [10], it was found that the world-wide airport network (WAN) featured the first case where $\langle C^w\rangle > \langle C\rangle$, indicating a *rich-club* phenomenon [20], whereas a scientific collaboration network [21] was found to have $\langle C^w\rangle \approx \langle C\rangle$. We observed that the weighted clustering coefficient of both RTS_D ($\langle C^w\rangle = 0.98$) and RTS_E ($\langle C^w\rangle = 0.97$) are only slightly higher than their topological analogs, implying that the weights on the RTS network are uncorrelated.

The BUS networks, however, have weighted clustering coefficients $\langle C^w\rangle = 0.66$ for BUS_D and $\langle C^w\rangle = 0.63$ for BUS_E, approximately 16% higher than their topological measures. As $\langle C^w\rangle > \langle C\rangle$, we can conclude that closed triangles are more likely formed by edges with larger weights. The BUS networks also appear to have similar average weighted clustering coefficients to those found for the German railway ($\langle C^w\rangle_{GR} = 0.75$) and the Indian railway ($\langle C^w\rangle_{IR} = 0.69$) [15].

For a better understanding of network cohesiveness, we can average C_i and C_i^w over all nodes with a certain degree k to yield the clustering spectrum $C(k) = \frac{1}{Np(k)} \sum_{i/k_i=k} C_i$ and weighted clustering spectrum $C^w(k) = \frac{1}{Np(k)} \sum_{i/k_i=k} C_i^w$. Both $C(k)$ and $C^w(k)$ appear to be independent of k; the clustering spectrums for the RTS networks (Fig. 7(a)) are fairly constant, with a narrow range ($(0.87, 1)$ as compared to $(0.2, 0.8)$ for inter-municipal traffic in Sardinia [14]).

In contrast, the clustering spectrums $C(k)$ and $C^w(k)$ for the BUS networks (Fig. 7(b)) follow the scaling law $C(k) \sim k^{-\beta}$ for $44 < k < 500$, implying that the majority of the nodes have a moderately hierarchical connectivity [22]. We also observed that nodes with high degree k have lower clustering coefficients than one would expect given the power-law fits. We found these high-degree nodes to be interchanges with buses going to different parts of Singapore. If we consider the BUS network to have a hierarchical star-like topology [22,8], it would be reasonable that the neighbors of interchanges would be less connected than lower-degree local neighborhood bus stops.

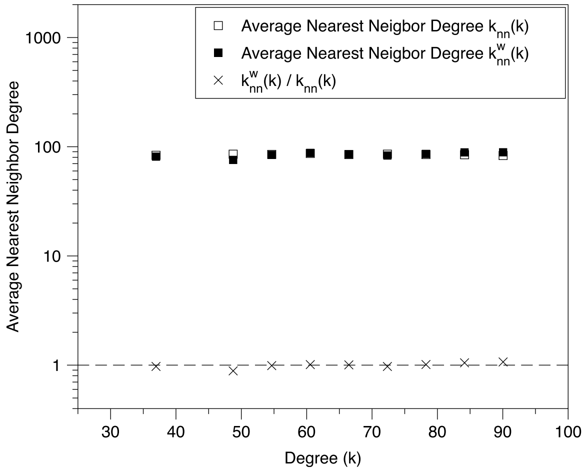
Comparing the weighted and topological clustering spectrums, we see that $C^w(k) > C(k)$ up to $k < 500$, indicating that, for this degree regime, clustering is formed by edges with higher weights. In fact, we observed that $C^w(k)/C(k)$ increased with k, suggesting that the more connected a bus stop, the more closed triangles were formed with high traffic routes, until a transition point at $k \approx 500$. From this point onwards, the trend dissipates, and no clear relationship between $C^w(k)$ and $C(k)$ is observed. A possible explanation for this discrepancy is that the interchanges also function as RTS stations, and travel between other geographically distant interchanges is more likely on the faster rail system.

3.4. Degree–degree correlations

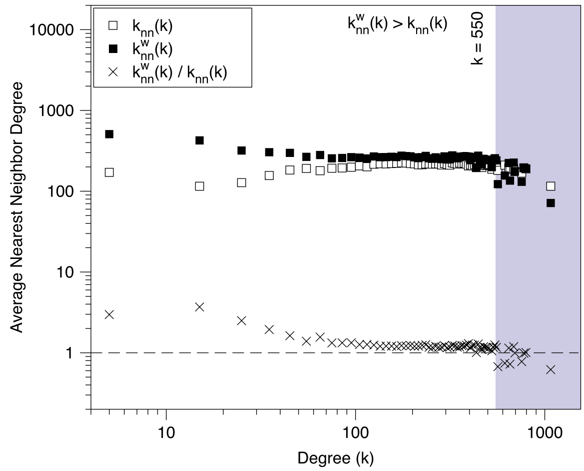
In this subsection, we examine another important topological characteristic of a network: the degree–degree correlation between connected nodes. We say that a given network is *assortative* if the high-degree nodes have a tendency to connect to other high-degree nodes. For *disassortative* networks, low-degree nodes tend to connect to high-degree nodes.

In 2002, Newman introduced a summary statistic for assortativity, r, defined as the Pearson correlation coefficient of the degrees at either ends of an edge [23]. The assortativity of a network r lies in the range $[-1, 1]$, where -1 indicates a completely disassortative network and 1 indicates a completely assortative network. An ER random graph has assortativity 0. Examples of assortative networks with $r > 0$ [23] include the network of film actors [2] and scientific co-authorships [24,25]. The WWW [3], neural network of *Elegans* [2] and freshwater food web at Little Rock Lake in Wisconsin [26] are disassortative networks, with $r < 0$ [23].

For both RTS_D and RTS_E, we observed a slightly negative topological assortativity, r, of -0.088 and -0.078, respectively, similar in magnitude to that of the Indian railway network [5], the WWW [3,23], and several of the city transportation



(a) RTS_D (weekday).



(b) BUS_D (weekday).

Fig. 8. Degree similarity spectrum for the RTS and BUS networks.

networks in Poland [8]. On the other hand, the assortativity of BUS_D and BUS_E is 0.015 and 0.06, respectively, indicating slightly assortative networks. That said, an assortativity of close to zero for all four networks implies that the tendencies are mild and the networks are largely egalitarian; nodes appear to connect without preference (at least from a topological perspective).

A closer examination of the four travel route networks can be performed using another measure, the *average nearest-neighbors degree*,

$$k_{nn,i} = \frac{1}{k_i}\sum_{j=1}^{N} a_{i,j}k_j, \tag{6}$$

which indicates if a node i's neighbors have a similar degree to that node i. As with the clustering coefficients, we can average $k_{nn,i}$ for nodes of a given degree to give the similarity spectrum $k_{nn}(k) = \frac{1}{Np(k)}\sum_{i/k_i=k} k_{nn,i}$. If we observe $k_{nn}(k)$ to increase with k, the network is *assortative*. If $k_{nn}(k)$ decreases with k, the network is *disassortative*.

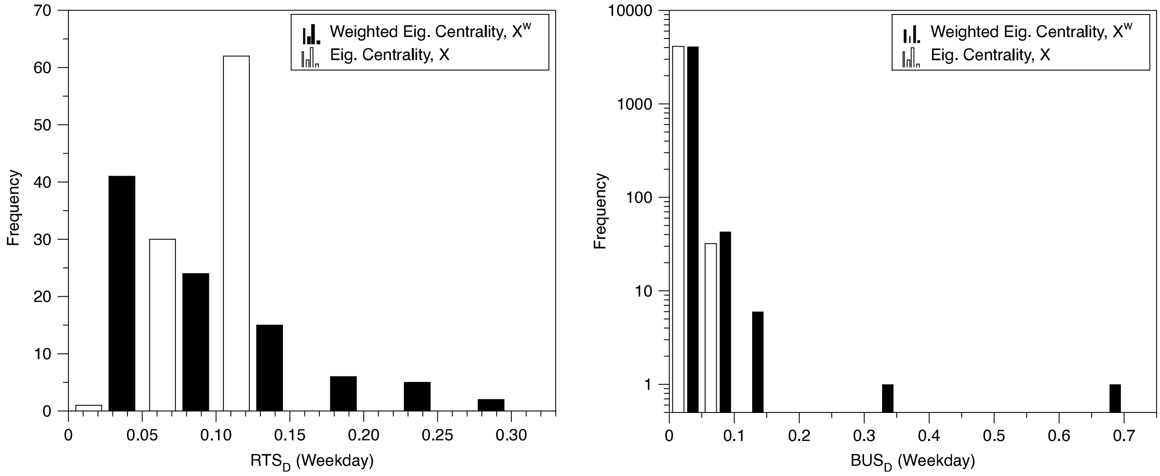
The *weighted average nearest-neighbors degree* is defined by

$$k^w_{nn,i} = \frac{1}{s_i}\sum_{j=1}^{N} a_{i,j}w_{i,j}k_j, \tag{7}$$

and it measures the *affinity* of nodes to connect to high-degree or low-degree neighbors depending on the edge weights. The weighted similarity spectrum $k^w_{nn}(k) = \frac{1}{Np(k)}\sum_{i/k_i=k} k^w_{nn,i}$ also indicates the assortativity of the network, but takes into the account the effect of weighted edges. If $k^w_{nn,i} \approx k_{nn,i}$, then the edge weights are uncorrelated with the degree of i's neighbors. However, if $k^w_{nn,i} > k_{nn,i}$, then heavily weighted edges connect to neighbors with larger degree, with the opposite occurring when $k^w_{nn,i} < k_{nn,i}$. In their 2007 study of the inter-municipal traffic of Sardinia [14], Montis et al. found that $k^w_{nn,i} > k_{nn,i}$, indicating that large municipalities exchanged a large number of commuters. Similar assortative behavior was found in the world-wide air transportation network and the network of scientific co-authorships [10].

Returning to the RTS networks, although a negative r indicated that the network was slightly disassortative, recall that the degree distributions in Section 3.2 showed that the RTS networks are highly connected clusters. In fact, the weekday similarity spectrum plotted in Fig. 8(a) (the weekend spectrum is similar) illustrate that the average nearest-neighbor degree is fairly constant, i.e., regardless of the degree of a given node, its neighbors have an a similar degree $k_{nn}(k) \approx \langle k \rangle$. This implies that the low-degree nodes connect primarily to the nodes in the main cluster, rather than other lower-degree nodes. The weighted similarity spectrum, k^w_{nn}, is similar to its topological counterpart (Fig. 8), with $k^w_{nn}/k_{nn} \approx 1$ for both weekday and weekend networks, signifying that the traffic is uncorrelated with the degree of an RTS station's neighbors.

For the BUS networks, Fig. 8(b) shows that the similarity spectrums of the unweighted networks are slightly assortative. However, the weighted degree similarity spectrums $k^w_{nn}(k)$ of the networks appear *disassortative*. Also, since $k^w_{nn}(k) > k_{nn}(k)$, at least up till $k \approx 500$, heavily weighted edges connect to neighbors with larger degree, indicating a larger flow of passengers per edge to more well-connected locations. The ratio between the weighted and unweighted degree similarities show that, unlike the clustering spectrums, $k^w_{nn}(k)/k_{nn}(k)$ decreases over k until $k \approx 500$, after which the relationship between $k^w_{nn}(k)$ and $k_{nn}(k)$ becomes less clear.



(a) RTS_D.

(b) BUS_D.

Fig. 9. Topological and dynamical centrality distribution of the RTS and BUS networks.

3.5. Eigenvector centrality

In Section 3.2, we described the degree and strength of a node as indications of its centrality or importance in a network. Degree and strength are the simplest measures of centrality and they make the assumption that the importance of node derives solely from the *quantity* of edges incident upon in (and the weights of those edges in the case of strength centrality). There exist other measures of node centrality such as *betweenness* and *closeness*, which are based on network paths. However, these measures are defined only for simple graphs without weighted edges.

In this study, we use the eigenvector centrality measure [27], which forms the foundation for the PageRank algorithm used by Google [28]. The basic concept underlying eigenvector centrality is that the "quality" of an edge should matter, i.e., an edge to a highly central node should matter more than an edge to a node with low centrality. As such, a node's centrality should depend on the centrality of its neighbors. If we let x_i be the centrality score for node i, then we can formalize this concept as follows:

$$x_i = \frac{1}{\lambda}\sum_{j=1}^{N} a_{i,j}x_j, \tag{8}$$

where λ is a constant value. Written in vector-matrix notation,

$$\lambda \mathbf{x} = \mathbf{A}\cdot\mathbf{x}, \tag{9}$$

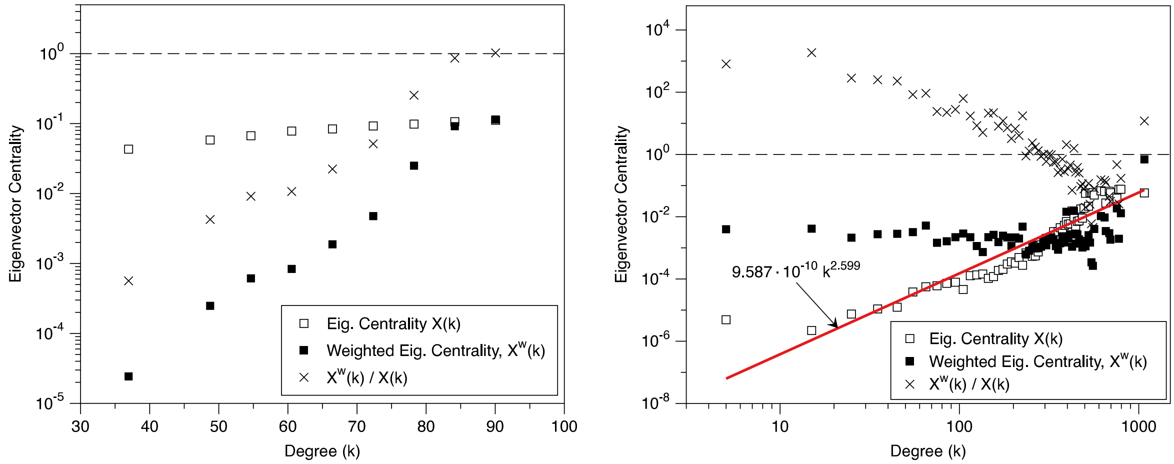
and hence $\mathbf{x}$ is an eigenvector of the adjacency matrix $\mathbf{A}$ with eigenvalue λ. Using the Perron–Frobenius theorem, we can show that λ is the largest eigenvalue and $\mathbf{x}$ is the associated eigenvector [29]. If we normalize $\mathbf{x}$, the eigenvector centrality of a node varies in the range (0, 1), with larger values indicating higher centrality. We can easily extend this concept to weighted networks by noting that weights should affect the importance of edges,

$$x_i^w = \frac{1}{\lambda}\sum_{j=1}^{N} a_{i,j}w_{i,j}x_j, \tag{10}$$

and following the same arguments as before. We denote the average centrality of a network as X and the average weighted centrality as X^w.

The topological and dynamical centrality distributions for the RTS networks are plotted in Fig. 9(a). The topological centralities in both RTS_D and RTS_E are relatively low, with $x_i \in (0.04, 0.11)$, with distributions that skew left. We observed that more than 50% of nodes have similar centrality scores; this is unsurprising, since the RTS network is almost fully connected. From a topological perspective, most nodes appear nearly equal in terms of importance. However, a dynamical perspective tells a different story; the distribution of x^w also differs significantly from the topological case, featuring a right skew with many nodes of little importance and a few hub nodes with high centrality. This result is compatible with our understanding of the RTS networks thus far: the traffic between nodes is scale-free.

The eigenvector centrality spectrums $X(k)$ and $X^w(k)$ are plotted in Fig. 10(a). It is clear that the topological eigenvector centralities appear quite constant, but the weighted centralities appear to increase linearly with degree k on the semi-log scale. Also, other than nodes with the largest degrees ($\approx$90), the weighted centralities are lower than their topological analogs. This can be explained by the fact that the number of passengers to a given node grows at a faster rate than the



(a) RTS_D (weekday).

(b) BUS_D (weekday).

Fig. 10. Centrality spectrum for the RTS and BUS networks.

number of connections it has. Hence, we would expect that, when accounting for edge weights, the centrality scores of the lower-degree nodes will be reduced.

Comparing the weekday and weekend eigenvector centralities of RTS stations, we can see how the importance of a given node changes over the week. Both the weekday and weekend topological centralities appear to follow a linear relationship (see Fig. 11). A similar linear correlation is observed for the weighted centralities, with the exception of two outliers: both outliers are within the central business district, which experience very high traffic during the weekdays but significantly lower traffic during the weekends, when many of the offices are closed. The station with the highest overall centrality services the main shopping district in Singapore, and it experiences high traffic during the entire week. In fact, we observed that the strength or centrality of a station is correlated with the amount of occupied shopping floorspace (Fig. 11(c)).

Both the weekday and weekend BUS networks have a low average centrality $\langle x \rangle \approx 0.01$, with a relatively small range $(0, 0.08)$. The weighted average centrality scores are a magnitude lower, with $\langle x \rangle \approx 0.003$, but with a wider range $x_i^w \in (0, 0.69)$. Fig. 9(b) shows the topological and weighted eigenvector centrality distribution for the nodes in the weekday and weekend BUS networks. We see that both distributions are heavily right skewed, but the weighted centrality distribution displays several very highly central nodes ($x_i^w > 0.40$). The node with the highest centrality is an interchange in the West of Singapore, attached to the last station[1] of the RTS line servicing that region. From the map in Fig. 1, it can be seen that there was a lack of RTS stations in that region as of 2008, making bus the primary mode of public transport. Interestingly, the stations that trail behind it in terms of centrality are not remarkable in terms of degree or strength. Their high centrality is due to their strong connection to the most central node, as they carry passengers from the interchange to surrounding areas (Fig. 12).

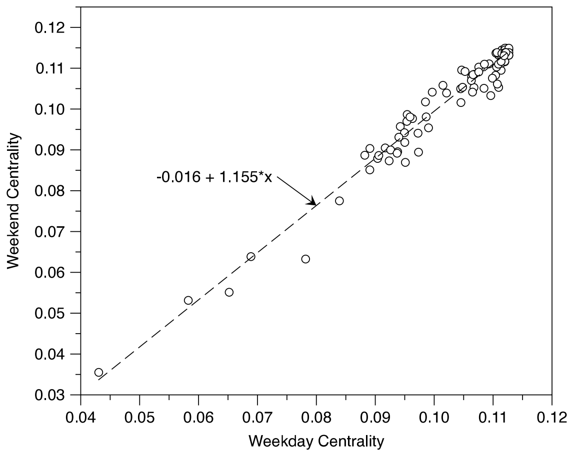
The BUS_D eigenvector centrality spectrums $X(k)$ and $X^w(k)$ are plotted in Fig. 10(b). The log–log plots clearly show that the eigenvector spectrum appears to increase linearly with k while the weighted eigenvector spectrum remains mostly constant, increasing slightly at large values of k. This constancy is unexpected, as nodes with low degree tend to also have low strength, as seen in Fig. 6(b), and would be expected to have low eigenvector centrality. A possible explanation is that nodes with low degree are connected mainly to nodes with high traffic (and thus, high importance), thus evening out the weighted centrality spectrum. This is supported by the disassortiveness of the weighted assortativity spectrum for low-degree nodes (see Fig. 8(b)).

4. Conclusions and future work

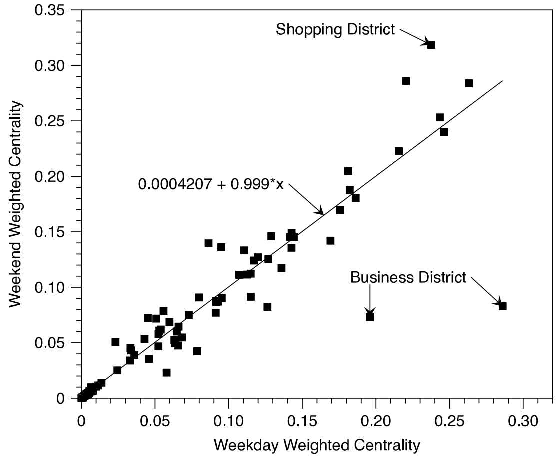
The desire to understand the nature of complex networks has formed the basis for a fascinating research field covering the span of many sciences. In this paper, we have analyzed the travel routes of the rail (RTS) and bus (BUS) public transportation systems in Singapore from a complex weighted networks perspective.

Our analysis shows that the dynamical properties of a network may differ significantly from its topological properties. In particular, the RTS network is topologically uninteresting; it is almost fully connected, and thus displays high clustering, with a small spectrum and almost neutral assortativity. From a dynamical (weighted) perspective, however, we observed that the traffic flows on the RTS network follow a power-law distribution, i.e., the network possesses hub nodes that experience very

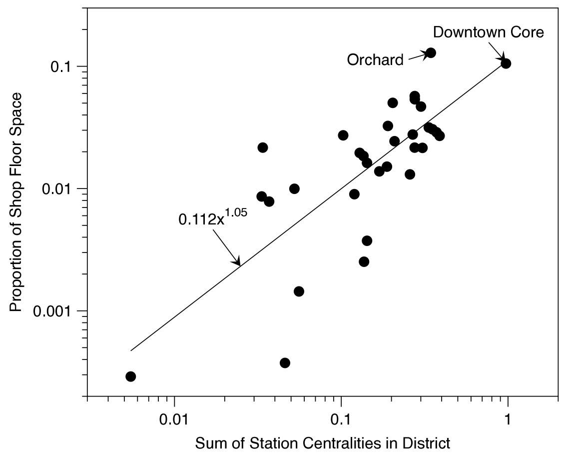
[1] An extension line of 3.8 km with two stations opened for revenue service on 28 February 2009. The extension line links the residential and industrial areas of that region to the last station of the existing RTS line, making it possible for commuters to have direct access to the RTS system.



(a) Topological centrality, X_i.



(b) Dynamical centrality, X_i^w.



(c) Shopping floor space plotted against centrality.

Fig. 11. Weekday and weekend centrality comparisons for the Singapore Rapid Transit System (RTS) networks.

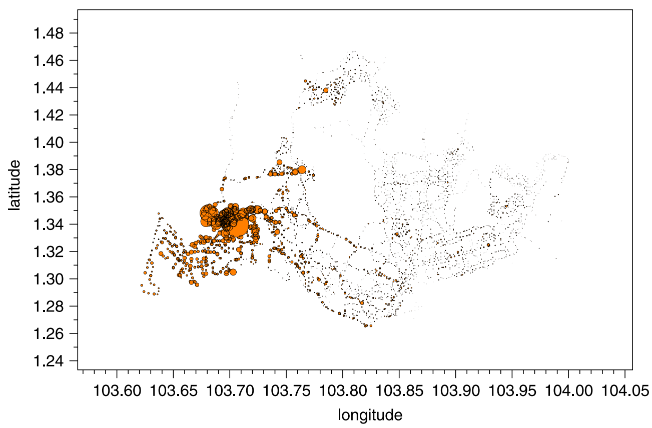


Fig. 12. The Singapore BUS network plotted geographically, with node sizes scaled by weekday weighted eigenvector centrality. The larger the nodes, the larger the centrality score.

high traffic. Moreover, although the topological centralities of the RTS nodes are equal during weekdays and weekends, we observe that the weighted centralities can differ significantly, particularly for nodes within the central business district.

The BUS weekday and weekend networks yielded their own interesting features. Unlike the RTS networks, which appeared to have strengths that decay exponentially, the strengths of the bus stops decay according to a power law, $p(k) \sim k^{-\gamma}$, indicating the presence of high-traffic hub nodes. The BUS network also appears to possess topological hierarchy, with a clustering spectrum that decays with degree k according to a power law, $C(k) \sim k^{-\beta}$. The weighted degree similarity spectrum adds to the picture, illustrating a slightly disassortative behavior.

Complex weighted analysis is a powerful tool for understanding a large complex system, and there are avenues for future work. A key limitation to this analysis was that we were only able to obtain data for a single week. It would be interesting to conduct a larger-scale study in which fluctuations across weeks or even years could be analyzed. Moreover, if more fine-scale data could be obtained, perhaps on an hourly basis, we could derive and compare the in and out statistics separately for the networks. This would give us important insights into the travel routes that occur in a single day.

Although we have restricted our study to the transportation domain, it would not be difficult to apply our analyses to other classes of network. In particular, this study illustrates the importance of *temporal dynamics*; the weekday–weekend networks illustrate pronounced differences (for example, in node centrality and weight distribution). Other real-world networks (e.g. gene networks, neural networks and the WWW) also experience changes in properties over time, and descriptions of real-world networks would be incomplete without taking these temporal changes into consideration. Given more data, future work would focus on quantifying these changes, for example, by examining trajectories on a phase-space plot and deriving appropriate models. Such analyses would provide richer information for complex network modelling and simulation, which is essential in a variety of research areas.

Acknowledgements

This work is supported in part by the Singapore A*STAR grant BMRC 06/1/21/19/457. The authors sincerely thank Dr. Stefan Ma, Dr. Gaoxi Xiao and Dr. Tomi Silander for their kind support and comments during this study.

References

[1] S.H. Strogatz, Exploring complex networks, Nature 410 (6825) (2001) 268–276.
[2] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[3] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[4] V. Latora, M. Marchiori, Is the Boston subway a small-world network? Physica A: Statistical Mechanics and its Applications 314 (1–4) (2002) 109–113.
[5] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, S.S. Manna, Small-world properties of the Indian railway network, Physical Review E 67 (3) (2003) 036106.
[6] W. Li, X. Cai, Empirical analysis of a scale-free railway network in China, Physica A: Statistical Mechanics and its Applications 382 (2) (2007) 693–703.
[7] L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Classes of small-world networks, Proceedings of the National Academy of Sciences 97 (21) (2000) 11149.
[8] J. Sienkiewicz, J. Hołyst, Statistical analysis of 22 public transport networks in Poland, Physical Review E 72 (4) (2005) 1–11. 10.
[9] G. Bagler, Analysis of the airport network of India as a complex weighted network, Physica A: Statistical Mechanics and its Applications (2008).
[10] A. Barrat, M. Barthélémy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proceeding of the National Academy of Science USA 101 (11) (2004) 3747–3752.
[11] J. Miguéns, J. Mendes, Weighted and directed network on traveling patterns, Bio-Inspired Computing and Communication (2008) 145–154.
[12] M.E.J. Newman, Analysis of weighted networks, Physical Review E 70 (5) (2004) 11.
[13] Onnela Jukka-Pekka, Saramäki Jari, Hyvönen Jörkki, Szabó Gábor, de Menezes M Argollo, Kaski Kimmo, Barabási Albert-László, Kertész János, Analysis of a large-scale weighted network of one-to-one human communication, New Journal of Physics 9 (6) (2007) 179.
[14] De Montis Andrea, Barthélémy Marc, Chessa Alessandro, Vespignani Alessandro, The structure of interurban traffic: a weighted network analysis, Environment and Planning B: Planning and Design 34 (5) (2007) 905–924.
[15] R. Praveen, A. Mukherjee, N. Ganguly, Hierarchical organization of railway networks, in: Proceedings of the European Conference on Complex Systems, ECCS-08, Jerusalem, Israel, 2008.
[16] Singapore Department of Statistics Singapore, Population and Land Area.
[17] International Monetary Fund, Gross domestic product based on purchasing-power-parity (PPP) per capita GDP.
[18] EZ-Link website, July 2009.
[19] G. Chowell, J.M Hyman, S. Eubank, C Castillo-Chavez, Scaling laws for the movement of people between locations in a large city, Physical Review E 68 (6) (2003) 1–7.
[20] S. Zhou, R.J. Mondragon, The rich-club phenomenon in the internet topology, IEEE Communications Letters 8 (3) (2004) 180–182.
[21] M.E. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications, Physical Review E, Statistical, Nonlinear, and Soft Matter Physics 64 (2 Pt 2) (2001) 026118.
[22] Ravasz Erzsébet, Barabási Albert-László, Hierarchical organization in complex networks, Physical Review E 67 (2) (2003) 1–7.
[23] M.E.J. Newman, Assortative mixing in networks, Physical Review Letters 89 (20) (2002) 208701.
[24] M.E.J. Newman, The structure of scientific collaboration networks, Proceeding of the National Academy of Science USA 98 (2) (2001) 404–409.
[25] Jerrold W. Grossman, Ion Patrick D.F., On a portion of the well-known collaboration graph, Congressus Numerantium 108 (1995) 129–131.
[26] Martinez Neo D., Artifacts or attributes? effects of resolution on the Little Rock Lake food web, Ecological Monographs 61 (4) (1991) 367–392.
[27] M.E.J. Newman, The mathematics of networks, The New Palgrave Encyclopedia of Economics (2007).
[28] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Computer networks and ISDN systems 30 (1–7) (1998) 107–117.
[29] Friedkin. Noah, Theoretical foundations for centrality measures, American Journal of Sociology 96 (1991) 1487–1504.