National University of Singapore

Revealing Spatial and Diurnal Dynamics of Spatial Interaction Networks – A Case Study of Singapore

GE6211 Project Report – Group P4

XU YUTING | A0113875B

4-4-2021

Contents

1. Introduction

Singapore is known for its long-term and comprehensive urban planning policies and effective implementation mechanism. Land use and transportation are planned in an integrated manner to serve the city-state's long-term development goals, and these range from economic competitiveness to a quality living environment to its people.

In the recent Master Plan 2019, greater emphasis has been placed on allowing greater flexibility in allowable land uses on specific land lots, such as encouraging amenity co-location, providing greater flexibility to landowners to repurpose industrial land in preferable configurations. In addition, to further spur economic growth and maintain its competitive advantages, the plan to build the Jurong Lake District as the second CBD has been announced, and a few new growth areas have been highlighted in the Master Plan, each specialising in respective industries. However, till now, we have limited understanding on how flexible land uses and new growth areas will influence the spatial and diurnal dynamics of human flows. Traditionally, the major employment areas are located in known locations such as the CBD, the western industrial areas, and other business parks/industrial estates that are scattered across the island. We do not know how such new emphasis will influence activity patterns and spatial interaction networks spatially and temporally, and whether these new planning directions will give rise to a different spatial organisation of the urban system.

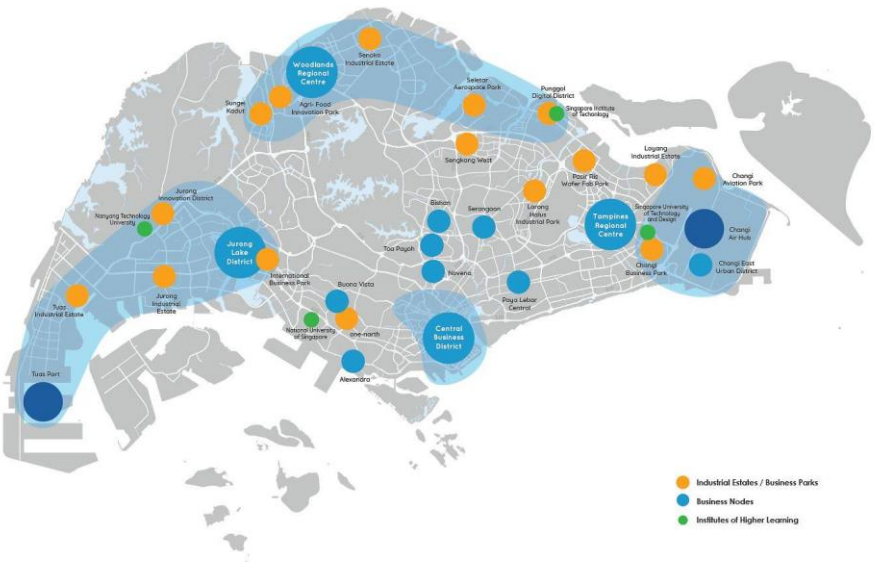


Figure 1-1 New Growth Areas highlighted in Master Plan 2019

An emerging discipline originated from statistical physics, network science has been increasingly used to study urban spaces, by conceptualising cities as a complex system, whose behaviours arise from its internal structure as a space of flows within networks. Such an approach allows us to gain new insights about the organisation of the urban space. A popular application of network science approach in the field of urban studies is with spatial interaction networks. A spatial interaction network is formed by commuting patterns that tell us about how places are functionally dependent. Through spatial interaction networks, a city can be conceived as a space of flows and interconnected places. Several past studies have been done on urban spatial interaction networks using public transport data to reveal urban spatial structure. For example, Zhong et al. (2014) used three years' public transport data (both bus and MRT) to identify the urban centres and hubs, and revealed community structure that resembles the authority's plan of a polycentric urban form. However, there are many classical network features and properties that remain unexplored but could facilitate our understanding of the present urban form.

Therefore, this project is positioned as an exploratory study to understand the spatial and diurnal changes in Singapore's urban spatial interaction networks. In particular, the study seeks to find out:

- Network Characterisation: Based on network topology and structure, what type of network is this?
- Spatial and temporal changes: How do spatial and temporal features of the network tell us about the activity patterns and urban spaces?
- Reflection of Techniques: How does uncertainty unfold in the data we are using and how it will be addressed?

The subsequent sections of the report will introduce the study area and the important information of Singapore's public transport network (*Section 2*), the datasets used (*Section 3*), the detailed methods and techniques employed in analysing the spatial interaction networks (*Section 4*), the results (*Section 5*), and key discussion points arising from the results (*Section 6*). The project will conclude with a reflection of the challenges encountered in this exploratory study and offer possible future research topics extending from this (*Section 7*).

2. Study Area

Singapore has an extensive and generally efficient multi-modal public transport system, comprising the Mass Rapid Transit (MRT), buses, and Light Rapid Transit (LRT) systems. The public transport network is also continuously expanded and improved to provide greater coverage and better services. As a result, public transport is the predominant transport mode for most residents and enjoys a high share of total ridership. As of 2019, average daily ridership for buses reached 4,099,000 passenger-trips, approximately more than 1 trip per resident per day (Land Transport Authority, 2020).

The design principle behind its extensive multi-modal system aims to differentiate the roles played by each mode of transport to achieve efficiency of transporting people across the island. In the "Hub-And-Spoke" model, the MRT serves as the backbone of the entire system by supporting long-haul travels and heavy transit corridors, but the buses and LRTs will serve as the feeder service that serve mostly local catchments and are primarily responsible for transporting people to the MRT stations (Centre for Liveable Cities & LTA, 2013).

Conventionally, the city centre of Singapore is recognised to be the Central Business District or the downtown core area in the south of Singapore, roughly encompassing Planning Areas such as Museum, Downtown Core, Singapore River and so on (Figure 2-1).

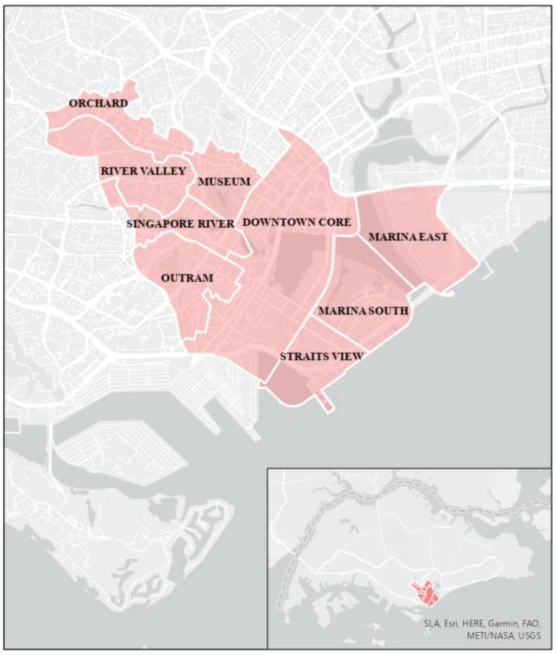


Figure 2-1 Conventionally recognised city centre of Singapore, mapped into Master Plan 2014 Planning Areas

3. Resources

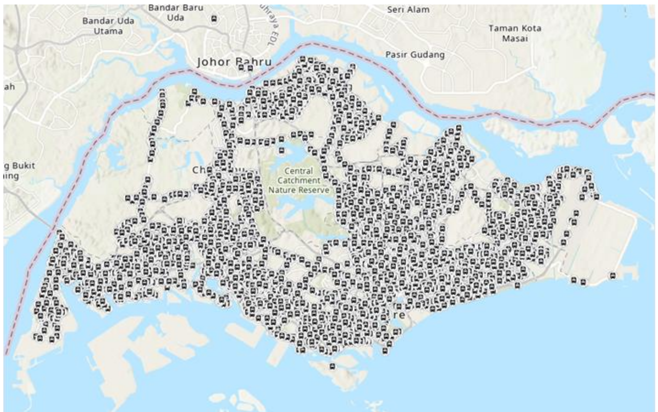
In this exploratory exercise, we used Passenger Volume by Bus Stop OD and Bus Stop Location data made available to the public via LTA[1] Data Mall.

The Passenger Volume By Bus Stop OD data was acquired for the month of January 2021 in .csv format. The dataset captures aggregated trip counts between each pair of Origin and Destination bus stops (denoted by Bus Stop Code) within the entire month, for each day type (weekday or weekend/holiday) and each hourly timeframe of the day. A sample of the data can be found in Figure 3-1.

YEAR_MONTH	DAY_TYPE	TIME_PER_HOUR	PT_TYPE	ORIGIN_PT_CODE	DESTINATION_PT_CODE	TOTAL_TRIPS
2020-12	WEEKEND	12	BUS	93201	92111	2
2020-12	WEEKDAY	12	BUS	93201	92111	5
2020-12	WEEKEND	23	BUS	59019	60179	1
2020-12	WEEKEND	13	BUS	1239	3059	1
2020-12	WEEKDAY	13	BUS	1239	3059	13

Figure 3-1 Sample Passenger Volume by Bus Stop OD Data

The Bus Stop Location data is acquired in shapefile format, last updated in December 2020. Each bus stop is identified by a bus stop name and Bus Stop Code (Figure 3-2). A total of 5040 bus stops are included in the dataset.



BUS_STOP_N	BUS_ROOF_N	LOC_DESC
63359	B01	HOUGANG SWIM CPLX
64141	B13	AFT JLN TELAWI
83139	B07	AFT JOO CHIAT PL
55231	B02	OPP SBST EAST DISTRICT
55351	B03	OPP FUDU WALK P/G
92089	B10	CHIJ KATONG CON

Figure 3-2 Bus Stop Locations in Singapore and sample dataset screenshot

[1] The Land Transport Authority (LTA) is a government agency responsible for the planning and implementation of land transport in Singapore.

4. Methods

This study employs a combination of network science techniques and spatial analysis methods to explore the spatiotemporal dynamics in the spatial interaction networks. Figure 4-1 presents the overview of the methodology used in this study.

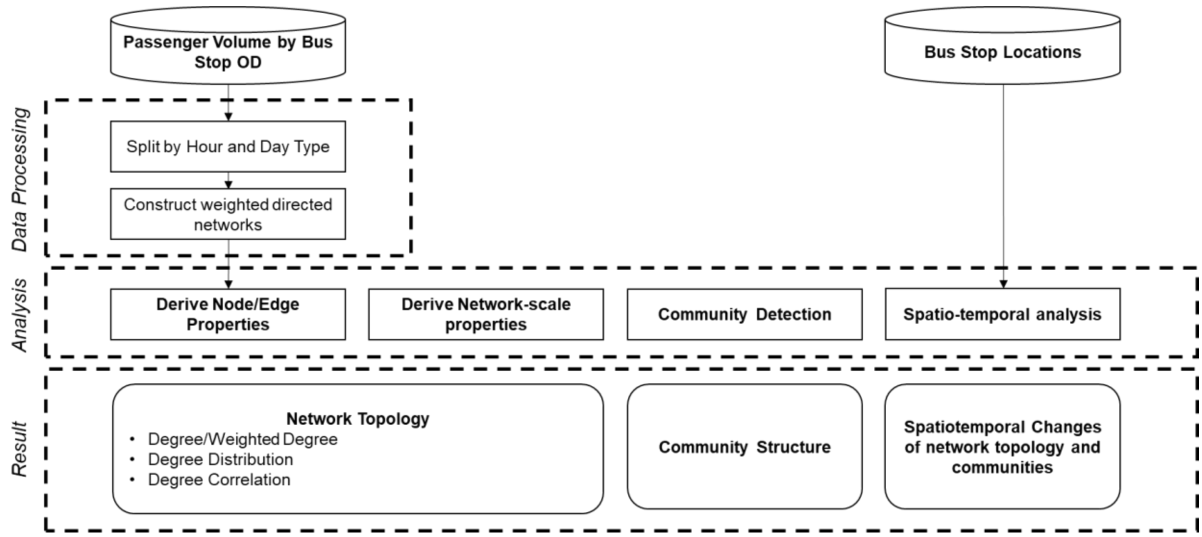


Figure 4-1 Overview of Methodology

4.1. Define and Construct Networks

The Passenger Volume data is first split into subsets of data organised by day type and hour of day. Each subset is then used to construct a weighted directed network, defined as G ≡ {V, E, W}, where:

- the vertices (V) of the network represent the bus stops, or the origin/destination of trips;
- the edges (E) represent the trips between each pair of origin/destination bus stops; and
- the weights (W) represent the trip volume/trip count for each pair of origin/destination bus stops.

This process of constructing networks is then repeated for each subset of the passenger volume data. The final output from this step will be 48 networks representing each hourly timeframe on a weekday and weekend/holiday. In a network, nodes that are linked by one edge are considered neighbours. It is worth noting that in the spatial interaction networks we constructed, there are two necessary conditions for an edge to exist between a pair of nodes. Firstly, there must be a bus service that serve the two bus stops. Secondly, there must be passengers who tap in/board the bus at one of the bus stops, and tap out/alight at the other. Therefore, the formation of such networks is shaped by both the existing transport infrastructure and services in place, and passengers' choice of trip origin and destination.

4.2. Degree and Weighted Degrees

For each node in spatial interaction networks, we computed the *degrees* and *weighted degrees*. Degree represents the number of links that a node has to the other node in a network. For a directed graph, the degree is also separated into *in-degree* (links coming in to the node) and *out-degree* (links coming out from the node). Weighted degree accounts for the weight of edges that connect each node, computed as the sum of all edge weights that link to the node. These measures allow us to compare nodes within the same network.

In addition, to describe the overall state of a network, we derived *average degrees* and *average weighted degrees*. Average degree is calculated as the number of edges over the number of nodes in a network. For a directed network, the average degree is the same as the average in-degree or average out-degree. Correspondingly, average weighted degree is the total sum of weighted edges over the number of nodes. These two measures allow us to understand network size and edge density.

4.3.Degree Distribution

Degree distribution is the probability distribution of a randomly selected in the network with degree k. Degree distribution is a key descriptor of network topology. A few typologies of networks have been previously established by scientists using degree distribution (Figure 4-2). A random network, also known as Erdős–Rényi network, is one where nodes in a network exhibit a randomly distributed probability of connecting to each other. Such networks typically exhibit a poison degree distribution, where the probability is the highest closer to a mean value. Another prominent type of networks is scale-free networks, where the degree distribution follows a power law. These two types of networks exhibit distinct topologies that influence flow patterns within the networks.

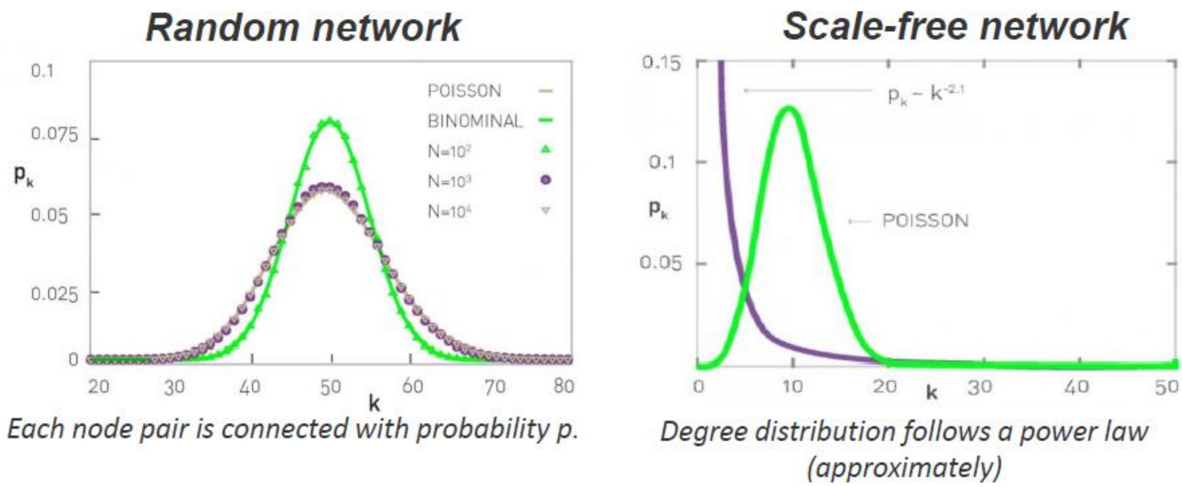


Figure 4-2 Network typologies established from degree distribution: (left) a random network exhibiting poison/binomial distribution of degree probability; (right) a scale-free network exhibiting power law distribution of degree probability. Illustrations from Barabasi (2016)

4.4.Degree Correlation

Degree correlation informs whether degrees in a network connect with nodes of similar or dissimilar degrees. There are two specific indices that can be used to quantify degree correlation: *Average Degrees of K-Nearest Neighbours of All Degree-K Node*, or *knn(k)*, calculates the average degree of all nearest-neighbours of nodes with degree k. *Joint Degree Correlation*, or JDD, is the probability of a randomly selected edge in the network being between nodes of degree i and j, illustrated by plotting the pair of node degrees for all edges. Using both knn(k) and JDD (Figure 4-3), we can categorise networks into whether they are:

a) Assortative, where high-degree nodes also tend to connect with high-degree nodes, thus producing a positive gradient in knn(k) plot and a positively inclined shape in JDD plot;
b) Neutral, where no distinct patterns of how high-degree nodes are connected, giving rise to an almost horizontal knn(k) graph and no distinct gradient in JDD plot; or
c) Disassortative, where high-degree nodes tend to connect with low-degree nodes, giving a negative gradient in knn(k) and JDD plots.

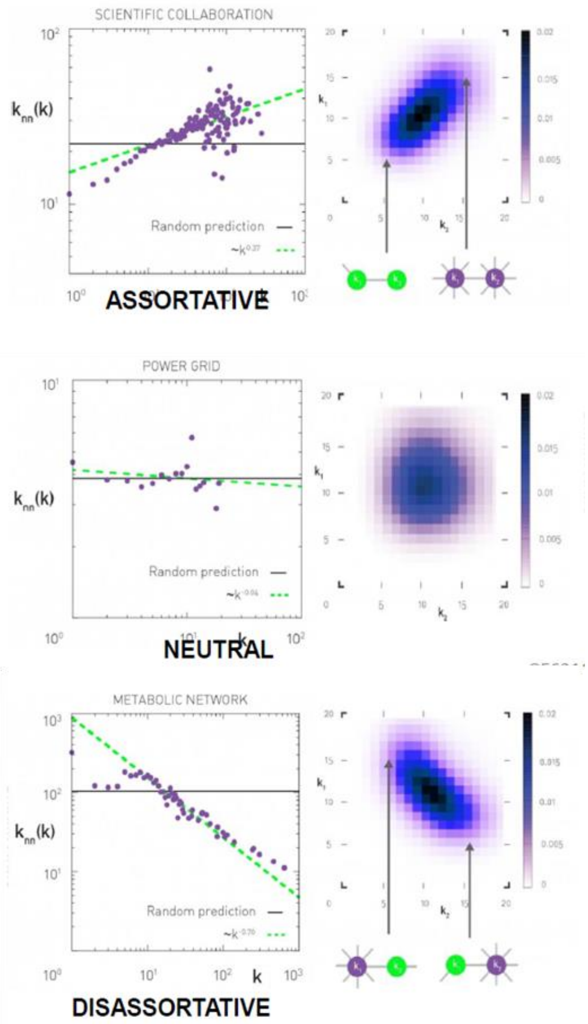


Figure 4-3 Expected network typologies based on Degree Correlation: (a) an assortative network; (b) a neutral network; and (c) a disassortative network. Illustration from Barabasi (2006).

In a spatial interaction networks, degree correlation is an important indicator of travel patterns. An assortative network would suggest that there is a high likelihood of people travelling between high-degree nodes to low-degree nodes, and therefore, functional hierarchies between these nodes, while a disassortative network would suggest otherwise.

4.5.PageRank Analysis

PageRank is a graph-based algorithm first used by Google Search to rank the search results. A node or a web page with higher PR if it is connected to other webpages that also have high PR, and ranked higher in the Google search results, because webpages with PR values are interpreted to represent quality content, high trustworthiness and hence, high importance in the web (Brin & Page, 1998). In urban network literature, PageRank is also interpreted as a measure of "centre-ness" in the urban form, such as in Zhong et al. (2014).

Based on Brin and Page (1998), PageRank is calculated as shown in *Equation 1*.

Equation 1 - PageRank

$$PR(A) = (1\text{-}d) + d\ (PR(T1)/C(T1) + \ldots + PR(Tn)/C(Tn))$$

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one.

Where:

- *d* is the dampening factor between 0 and 1;
- *C(A)* is the number of links going out from page A;
- *Ti* is a page that points to page A.

Since PageRank is a probability distribution for each network, PageRank values can only be used for comparison within the same network, and cannot be used to compare those calculated from other networks.

4.6.Community Detection

Community detection is a technique to identify nodes within the network that are more likely to be connected based on network topology. Depending on the specific techniques used, in a spatial interaction network, the communities could represent clusters of bus transit nodes where people are more likely to travel to.

There are many types of community detection algorithms and each deploys a different definition of community. We employed the Modularity algorithm (Blondel et al. 2008). In this algorithm, individual node will be iteratively assigned to each community and assessed using a quality function Q *(Equation 2)* to determine the best community structure in the network.

Equation 2 - Quality function Q for Modularity Community Detection Algorithm (Blondel et al. 2008)

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where

- A_{ij} represents the edge weight between nodes i and j;
- k_i and k_j are the sum of the weights of the edges attached to nodes i and j, respectively;
- m is the sum of all of the edge weights in the graph;
- c_i and c_j are the communities of the nodes; and
- δ is Kronecker delta function ($\delta(x, y) = 1$ if $x = y$, 0 otherwise).

The Modularity algorithm will identify non-overlapping communities where each node is assigned to only one community that best optimises the quality function. Each node starts as its own community, and modularity score is calculated for the initial state. Subsequently, each node is assigned to a neighbouring node to form another community, and the modularity score is computed. The process then repeats iteratively until the modularity score is maximised, and the node is eventually assigned to this community. The same process repeats again by merging the communities until the modularity score cannot be maximised further, and the final output will be non-overlapping communities.

4.7.Spatial and Temporal Analysis

Upon deriving the network properties, we analysed the spatial and temporal variations of such network properties, and visualisede the results in ArcGIS Pro. We used a simple framework below to guide our spatiotemporal analysis.

Table 4-1 Spatiotemporal analysis framework

		Spatial Dimension	
		Fixed	*Varied*
Time Dimension	Fixed	-	Within the same timeframe, do we observe spatial variations of network properties?
	Varied	At the same place, do we observe temporal variations in network properties?	Do the temporal variations vary with location? Do the spatial variations vary with time?

5. Result

5.1.Diurnal Lifecycle

Line plots of the average degree and average weighted degree of the entire network (Figure 5-1) reveals a general diurnal lifecycle of network evolution. The spatial interaction network undergoes broadly four stages following the expected urban activity intensity of the day. In the birth/growth stage (05:00 – 08:00, close to morning commuting hours), as more people start taking buses, the average degrees will increase rapidly (places become more connected), and averaged weighted degree will also increase rapidly (volume of trips increase). The trend then enters a stage of gentle growth or stagnation, where the average weighted degree still shows huge fluctuation that resemble commuting patterns, and starts contracting at around 19:00. The network eventually breaks down into multiple components after midnight.

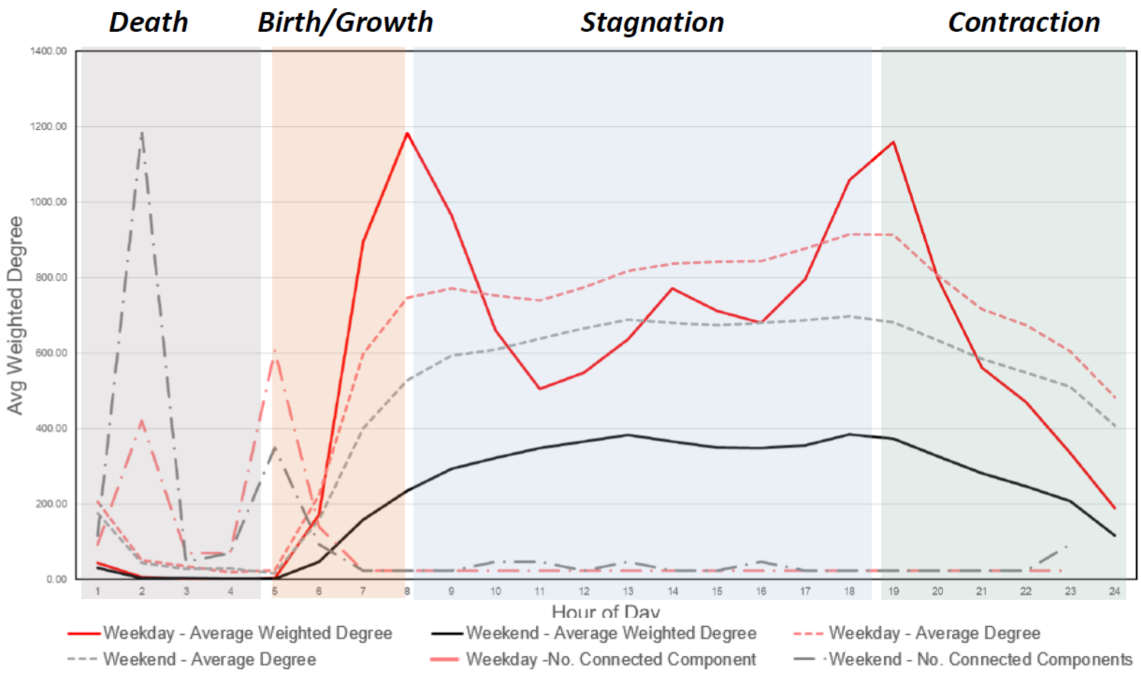


Figure 5-1 Diurnal Changes of Average Degree and Average Weighted Degree

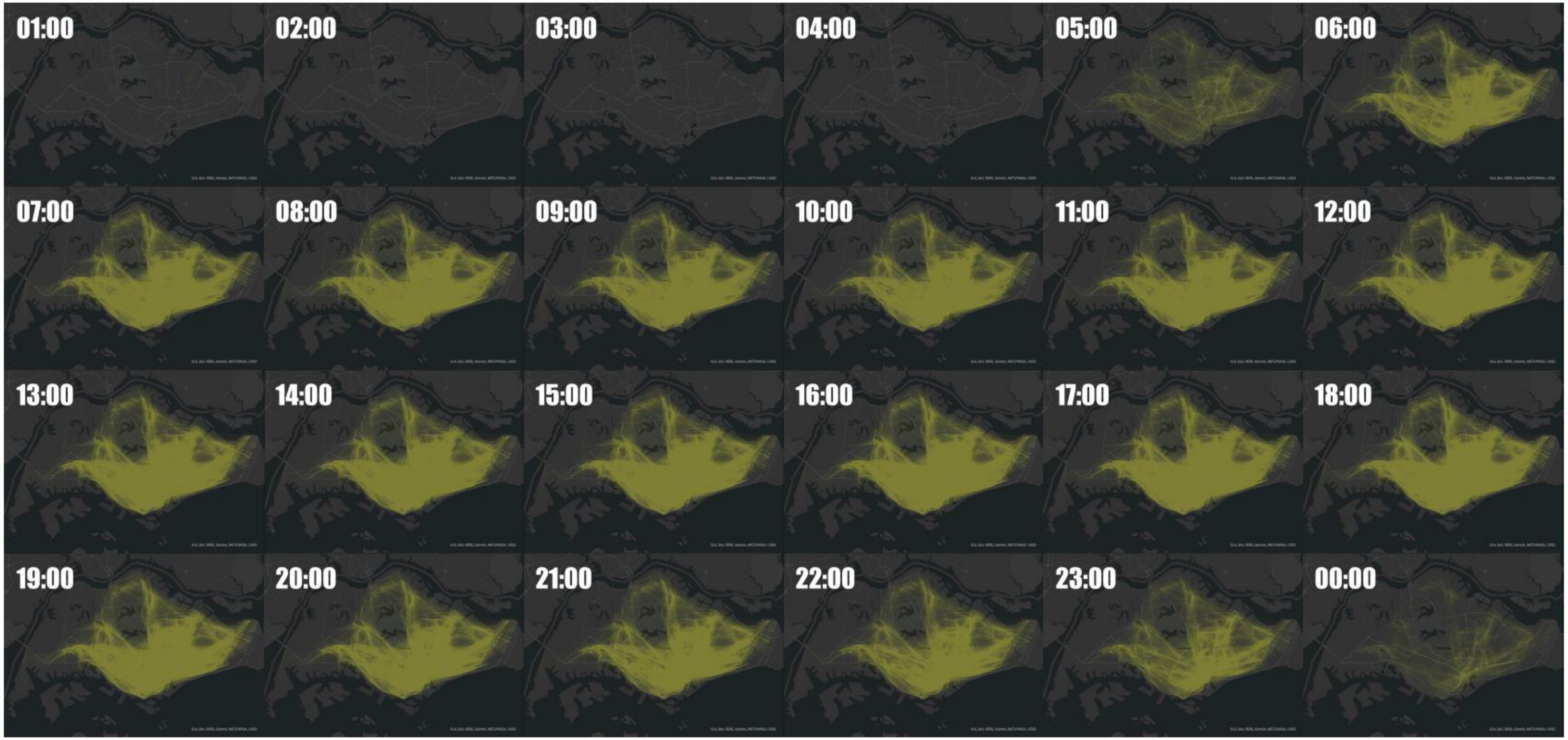


Figure 5-2 Mapping of network edges on a weekday in all hourly timeframes shows a general diurnal lifecycle where edge density increases in morning commuting hours, and gradually decreases in evening commuting hours.

Figure 5-3 Mapping of trip O-D pairs on a weekday 04:00 shows a disconnected network.

Similar observation of a general diurnal lifecycle can be made on the mapping of network edges by connecting the Origin and Destination bus stops with a line feature (Figure 5-2). Edge density undergoes visible increase from 05:00, roughly the time when most public transport services commence, reaches high density at around 09:00, followed by a period of stagnation or gentle growth. Edge density starts to decrease again after 19:00, around the evening commuting hours. An example of a disconnected network in the early mornings are shown in Figure 5-3, where we observe disjointed trip lines.

5.2.Spatial and Temporal Variations of Degrees and Weighted Degrees

Considerable spatial variations of degree differences were observed in every hourly timeframe. Ten transit nodes that were consistently among the top 10 highest degree nodes from 07:00 to 23:00 (periods of intense urban activities) on a weekday were identified (Figure 5-4a). It was found that all of these transit nodes are bus interchange or bus terminals. In Singapore, bus interchanges and bus terminals function as the start and end point for all bus services. Therefore, it is expected that these transit nodes will be serving more bus services and hosting higher passenger volumes than other bus stops. It is also worth noting that on a weekday, the top 10 bus stops with highest degrees in every hourly timeframe between 07:00 and 23:00 remain relatively consistent. This suggests that these bus interchanges consistently serve trips to and from a large number of places islandwide throughout the day, as well as relatively consistent trip patterns that allow these bus interchanges to be always of high-degree over the course of a weekday.

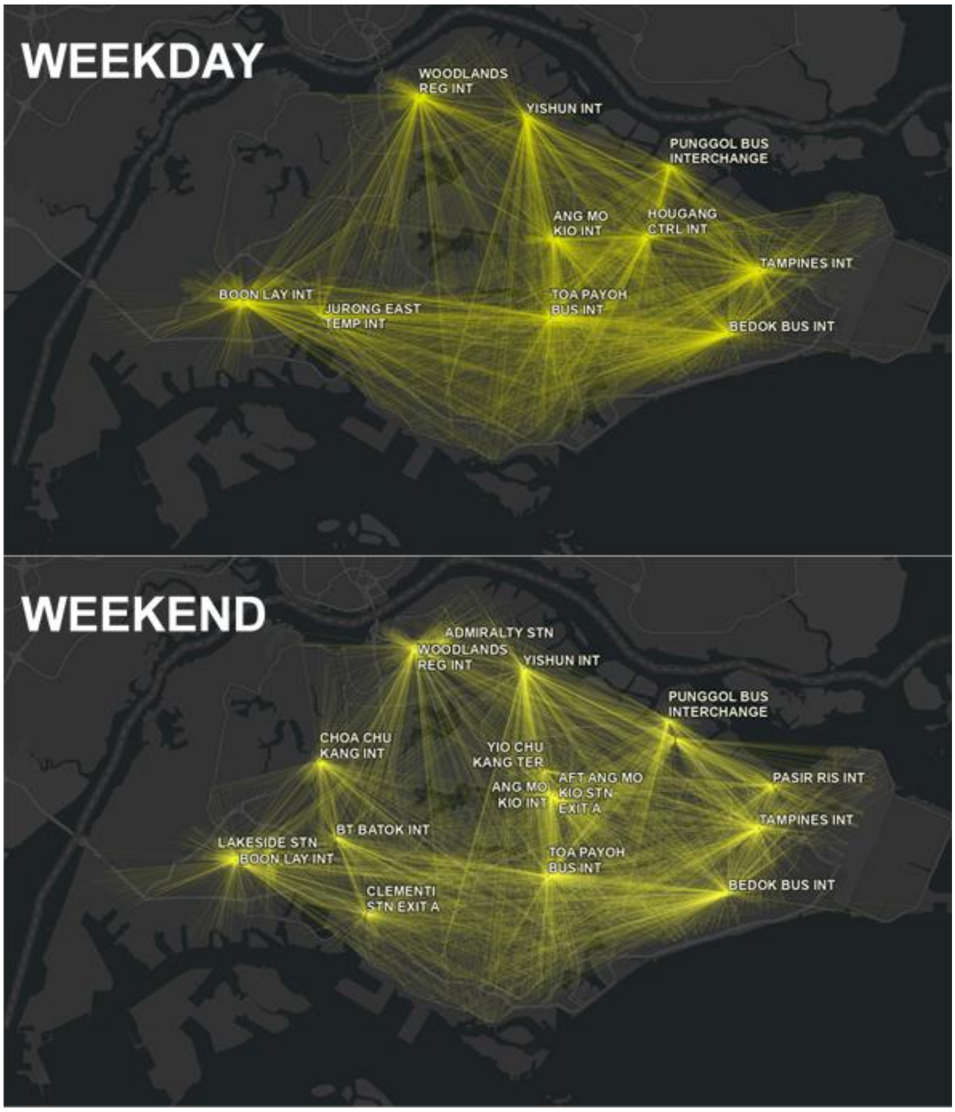


Figure 5-4 Nodes with the highest degrees from 07:00 to 23:00 on (a) a weekday and (b) a weekend.

Comparing high-degree nodes between a weekday and a weekend, we found that, over the same duration of 07:00 – 23:00, more bus stops made it to the top ten hourly highest-degree nodes, including several non-interchanges/non-terminals, such as the Lakeside MRT Station, Clementi Station Exit A, and Admiralty Station. These nodes, although do not serve as many bus lines as the bus interchanges, may still be preferred trip start/end location shown by their large degrees. It goes to suggest less temporally homogenous trip patterns on a weekend, possibly because of more diverse trip purposes on a weekend.

Within a weekday, we observed distinct temporal "pulses" of how weighted degrees change at respective locations (Figure 5-5). In particular, we could visibly identify two types of transit nodes among nodes with high weighted degrees:

1) Those that consistently have higher weighted degrees than other bus stops, such as Woodlands Regional Interchanges and Boon Lay Interchanges.
2) Those that experience periodical short "bursts" of high weighted degrees and subside rapidly within 1-2 hours, such as Lakeside Station, Admiralty Station.

It was also observed that these two types of bus stops tend to locate close to each other (e.g. Admiralty Station and Woodlands Regional Interchanges, Boon Lay Interchanges and Lakeside Station). It is thus intriguing to recognise two distinct patterns of weighted degree evolutions for bus stops that are located relatively close. Further research could look into how such patterns may correlate to local properties such as surrounding land uses/developments, and the commuting crowds they are serving.

Another important observation is the absence of prominent transit nodes in the conventionally recognised city centre/downtown areas of Singapore, such as the City Hall and the Marina Bay area. In fact, among the handful of prominent high-degree and high-weighted degree transit nodes we have observed so far, none of them is in the city centre.

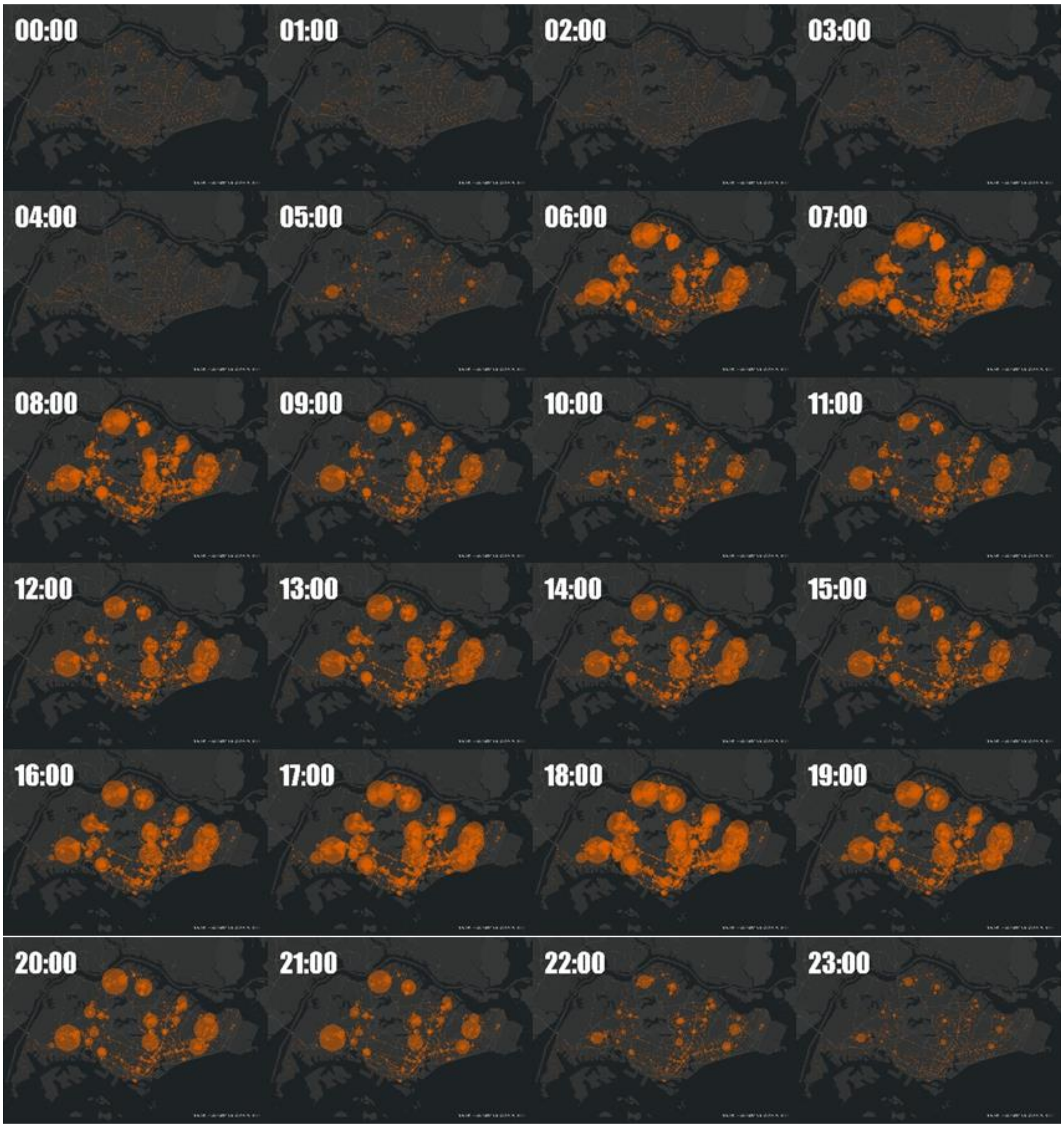


Figure 5-5 Hourly plot of Weighted Degrees on a Weekday

5.3. Degree Distribution and Scale-Free Networks

On both weekdays and weekends, we found that node degrees follow a long-tailed distribution that approximately resembles a power law (Figure 5-6(a)). In a power law distribution, there are many more low-degree nodes than high-degree nodes. Taking the logarithmic values on both axis in Figure 5-6(b), we show that the power exponent (gradient of linear regression) ranges between -0.84 to -0.91.

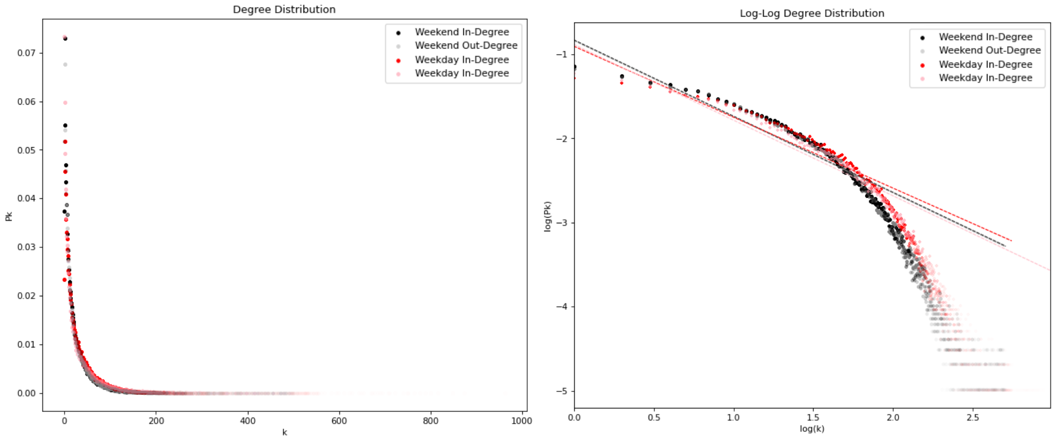


Figure 5-6 Degree distribution of overall weekday and weekend in/out degrees: a) probability distribution of degrees; b) logarithmic probability distribution of degrees.

The hourly plots in Figure 5-7 shows remarkable consistency in the degree distribution that similarly resembles a power law. Except for hours of between 01:00 to 05:00 when the networks are not fully connected, in the other hours, the power exponent/gradient of the linear regression of the logarithmic plot fluctuates only slightly around -0.9 (Figure 5-8). This means that for most hours of the day, there are only a handful of transit nodes that serve trips to and from many locations islandwide, while most transit nodes are only serving few locations. These transit nodes with highest degrees then possess disproportionate level of importance in the entire network. Networks with degree distribution that follows a power law are also known as scale-free networks (Barabasi & Albert, 1999). The important implications of this observation in the context of a spatial interaction network will be discussed in the next section.

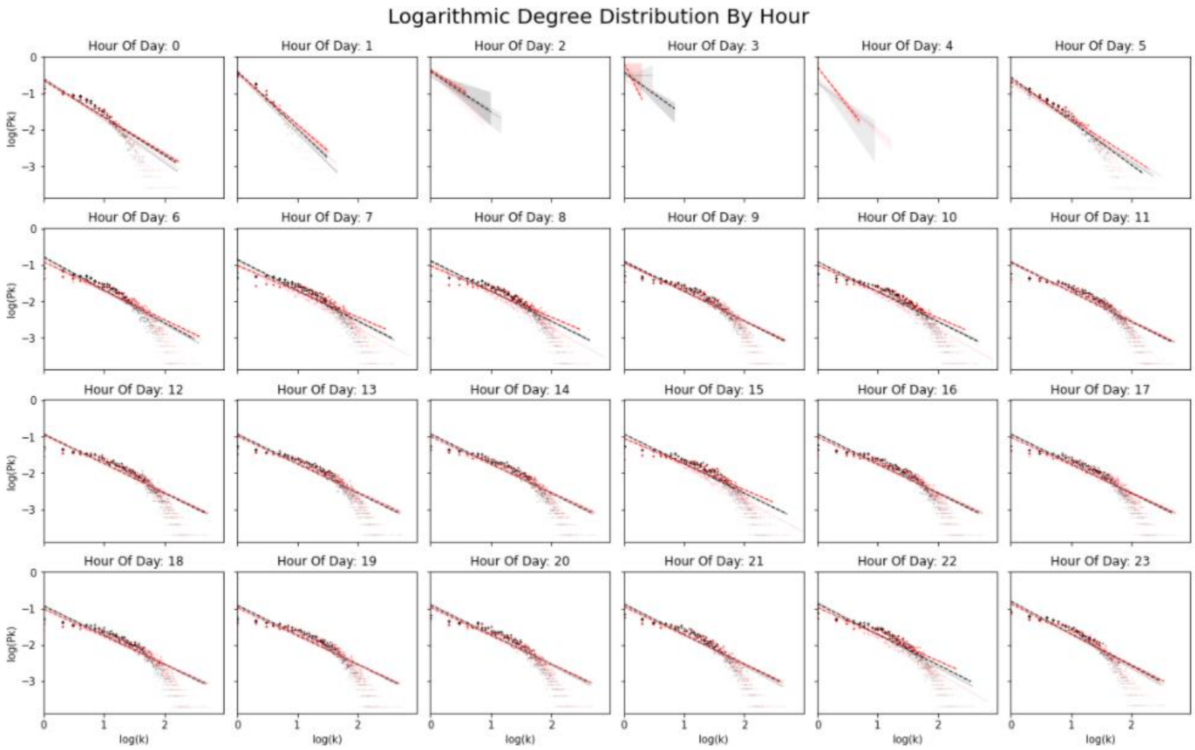


Figure 5-7 Hourly plot of logarithmic degree distribution

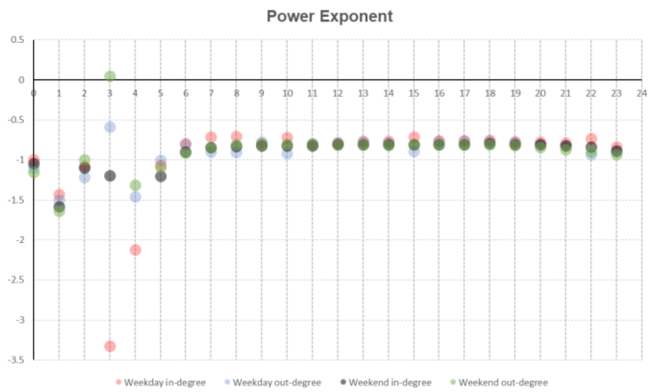


Figure 5-8 Power exponent of linear regression of hourly logarithmic degree distribution on a weekday

5.4.Degree Correlation

Two measures for degree correlations generated slightly contradictory results. In Figure 5-9, logarithmic knn(k) produces consistent hourly negative gradients, suggesting disassortative nature of the spatial interaction networks, where lower-degree nodes tend to have neigbouring nodes of higher degrees, while the higher-degree nodes tend to have neighbouring nodes of lower degrees. We also noticed that the range of log(knn(k)) is much smaller than that of log(k) in all hours. This means that, for each degree-k node, there might be a mixture of low- to high-degree neighbours that brings the overall log(knn(k)) to a smaller range than log(k) itself.

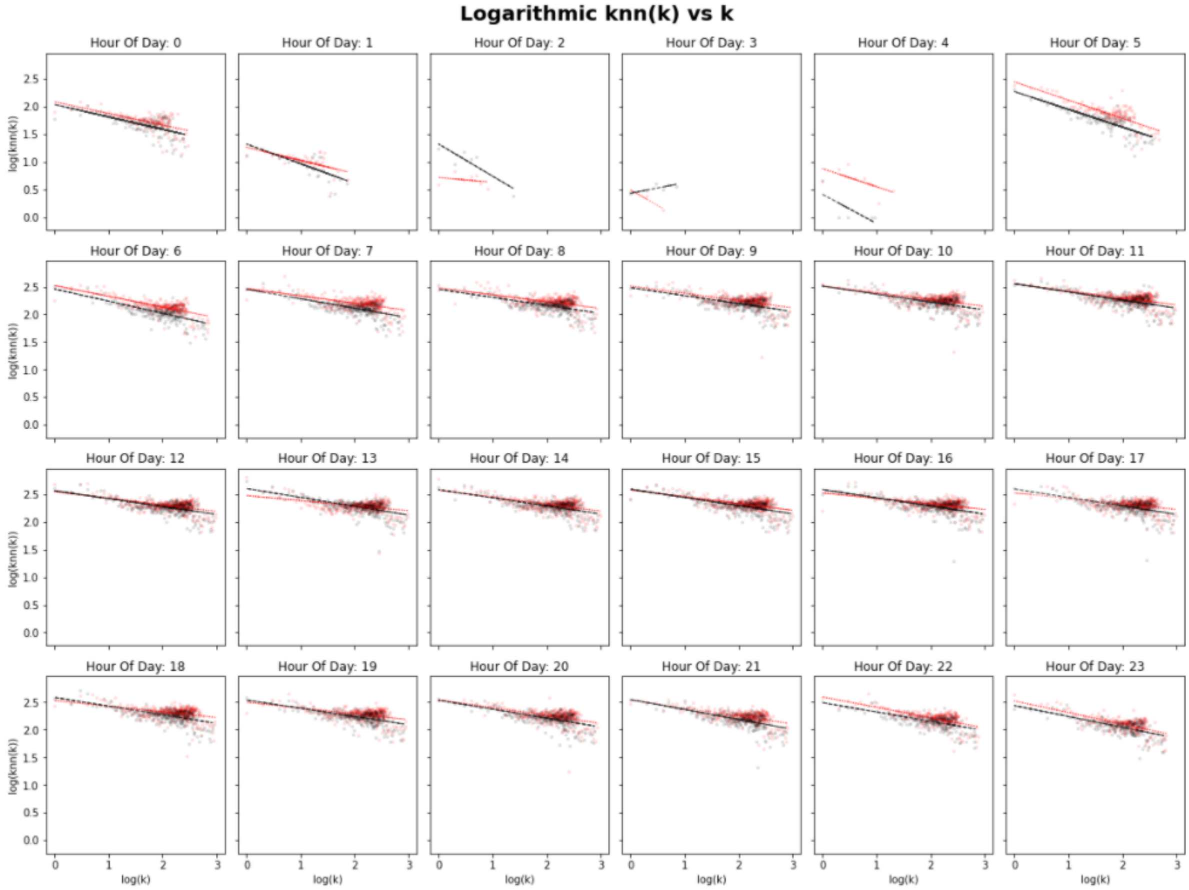


Figure 5-9 Average Degree of Nearest Neighbours of All Degree-k Nodes, or knn(k), shows a consistently negative gradient, suggesting a disassortative nature of the spatial interaction networks.

However, the finding is inconsistent with Joint Degree Distribution, another measure of degree correlation. From Figure 5-10, we observed neutral to positive correlation of neighbouring degrees, with shapes of density plot closer to that of a neutral network in Figure 4-3b). This suggest that, in any pair of neighbouring nodes, the neighbouring node tend to be of similar or slightly higher degrees. From JDD we also noticed a large distribution of neighbouring node degrees, which may have explained the smaller range of average degrees of neighbouring nodes observed in Figure 5-9. Since the results from two measures are inconsistent, further statistical tests would be required to ascertain the nature of the networks.

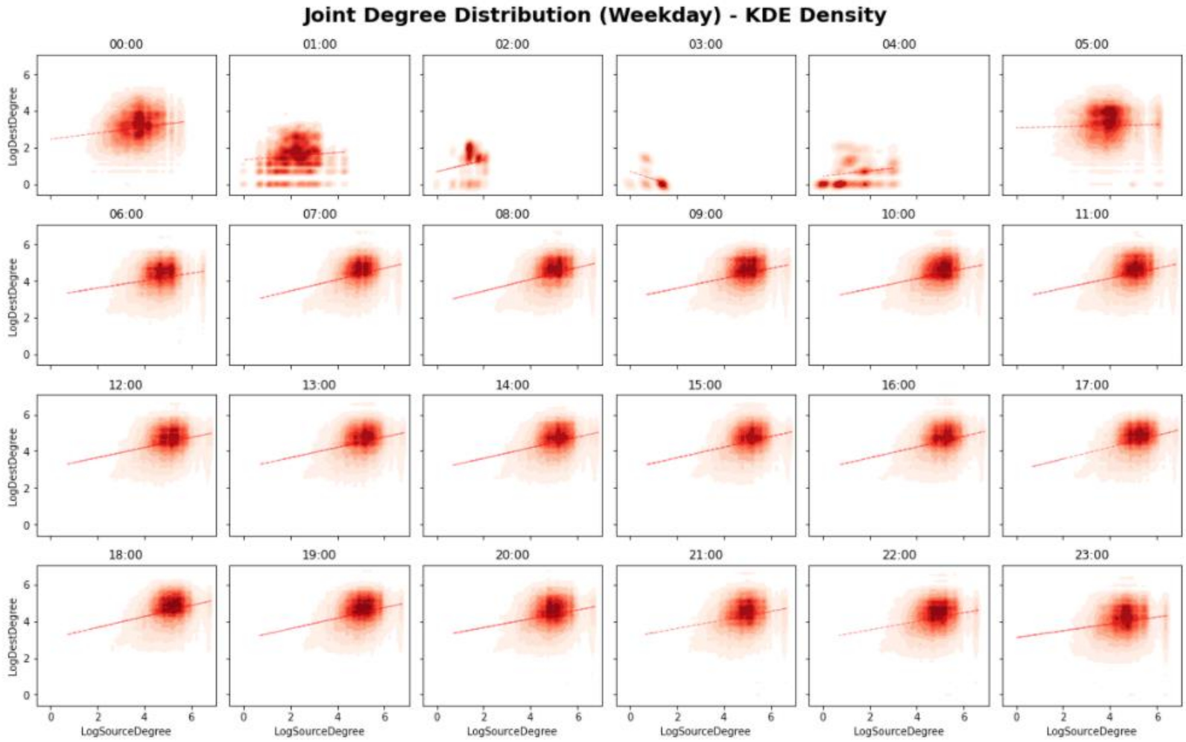


Figure 5-10 Joint Degree Distribution on weekdays

5.5.PageRank and Urban Centres

PageRank measures the relative "importance" of nodes in a network, and in many geospatial studies, transit nodes with high PR values are interpreted to be the urban centres. A few observations are made from Figure 5-11.

Firstly, we start to observe the conventionally recognised CBD and downtown areas through PageRank measures, which were not previously captured by degrees or weighted degrees. This could be understood from the definitions of the indices used. Degrees and Weighted Degrees are computed based on local properties: a node has high degrees if it is connected to many other nodes, and high weighted degrees if these links carry large weights. In contrast, PageRank emphasises network properties, or the quality of links a node has. Nodes will have high PageRank values if they are linked to other nodes with high PageRank values. In other words, the PageRank value reflects a node's connectivity to other important nodes, not just because the node itself has high degrees. The implications of this will be discussed further in the next section.

Secondly, within each hourly timeframe, distinct spatial variations in PageRank exist. This may be an obvious finding, but it illustrates quantitatively that places differ from one another in their centre-ness in the network. In some timeframes, there is one distinct area with high PageRank values (e.g. 06:00 and 11:00), in a commonly known urban form of monocentres. In others, a few areas may show similarly high PageRank values (e.g. 07:00, 10:00, 22:00), or polycentres.

Lastly, the urban "centres" determined using PageRank show considerable temporal variations. This reiterates our previous findings that, in different hours of the day, due to changing trip purpose and patterns, the importance of places in the entire city vary accordingly. For example, the conventionally recognised CBD and downtown areas show high PageRank values in a few time periods: 07:00 – 12:00, 15:00 and 22:00. In other time periods, Woodlands, Jurong East, and Tampines also show high PageRank values. These areas are also the key business nodes that were outlined in Master Plan 2019 (Figure 1-1) and are also popularly recognised as regional centres for commercial activities. What was perhaps surprising is the Seng Kang – Punggol stretch that also appeared to have high PageRank values in a few timeframes (07:00, 10:00, 15:00, 22:00), although not typically known to be conventional commercial centres but more for residential and industrial activities.

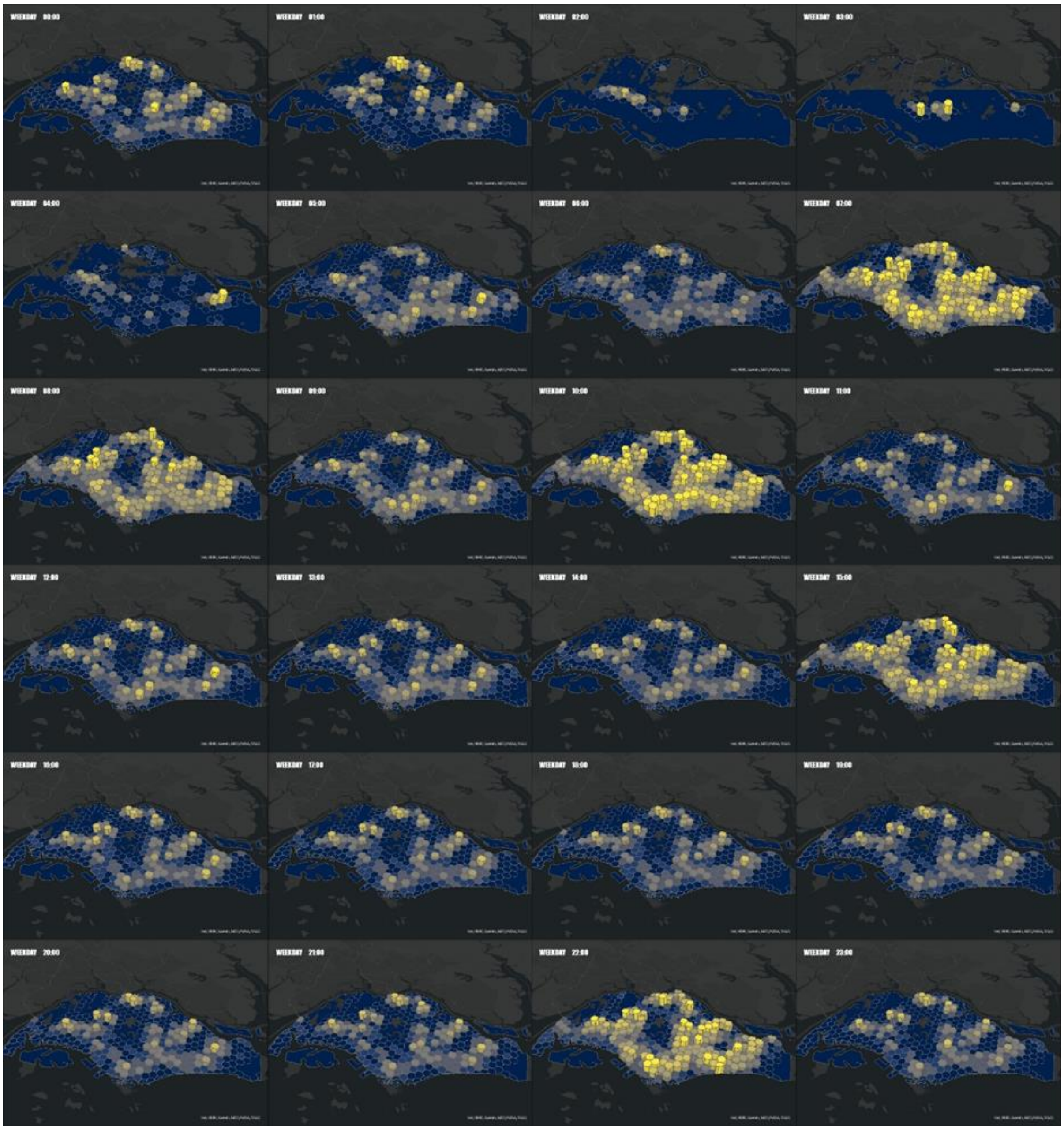


Figure 5-11 3D visualisation of PageRank kernel density on a weekday

5.6. Temporal Variations of Community Detection

Modularity algorithm is applied on the networks from all hourly timeframes on a weekday. We compared the results from several time slices with high trip volumes, such as 08:00 and 18:00 for commuting crowds, and 12:00 for lunchtime crowd. With these three time slices, we observed changing community memberships highlighted in the specific areas in Figure 5-12. It means that, over the course of a day, the network topology at these locations could be different, suggesting different travel patterns (choice of origin/destination) at these hours.

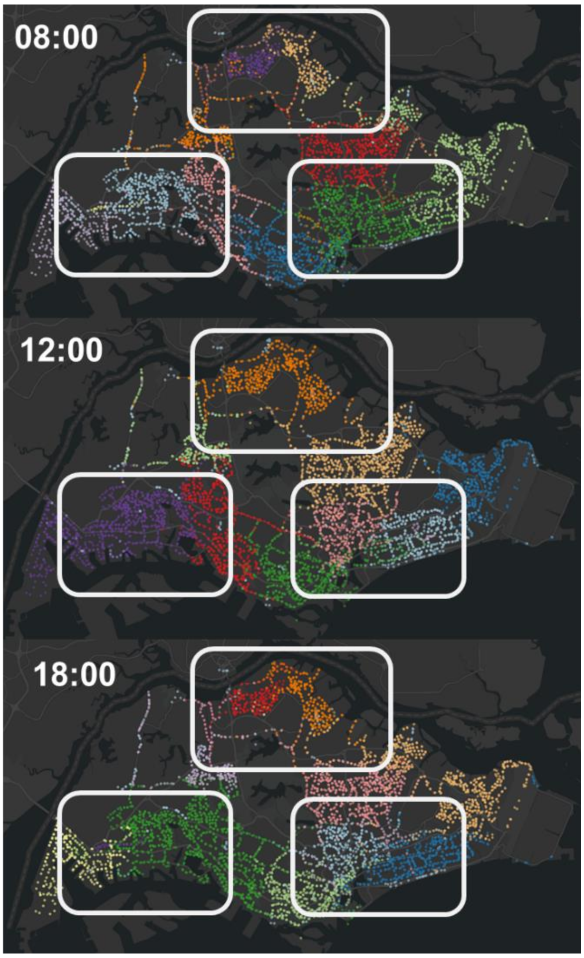


Figure 5-12 Community detection results on a weekday shows changes in membership structure

6. Discussion

In this section, we extend the key findings from Section 5 and discuss the implications of these findings in understanding urban space in Singapore.

6.1.Local vs Network Properties:

In Section 5 we illustrated that most of the nodes with high degrees/weighted degrees are usually transit hubs, which are understandable for the sheer number of bus lines they serve. We understand these to be local properties: interventions or actions taken at local scale/specific places to build a bus interchange or a bus terminal lead to the resultant high degrees seen at these transit nodes.

Figure 6-1 Comparison of PageRank, Weighted Degree and Degree of on Weekday 09:00

In contrast, nodes with high PageRank values are not necessarily transit nodes. Instead, PageRank captures areas known to be the CBD or downtown core, where there are no bus interchanges. By definition, PageRank does not measure local properties such as degrees; it measures “network properties” of a node by accounting for the quality of the neighbours that a node is linked to. A node is considered “important” with high PageRank values if its neighbours are also important with high PageRank, and this is beyond the local properties/what development takes place at the node location, but more on its topological characteristics in a network.

Since PageRank can capture the conventional city centres, it allows us to consider what defines a city centre, and more broadly, urban spatial structure. Zhang et al. (2021) reviewed definitions of urban spatial structure in existing urban study literature and broadly classified them into two classes: the morphological definition, which looks at spatial concentration of employment opportunities and activities, versus the functional definition, which is based on network topology. In Singapore’s case, we demonstrated with empirical evidence that the importance of a city centre lies in its network topology, or its connectivity to other important nodes in the city. Further research could look at whether two definitions provide different results of urban spatial structure to enhance our understanding of an urban system.

6.2.Local Rules to Global Emergence

This study uses individual trip data to study collective patterns that emerge from the entire urban system. The passenger volume data captures individual actions which reflect the outcome of a wide range of considerations from individual trips, such as origin and destination of the bus stops, timing of trips, exact services to take and where to make transfers. These considerations are also inevitably shaped by existing public transport infrastructure and services. Despite the wide range of considerations at the individual's level, the collective trip data revealed important emergent patterns at the system's level with remarkable consistency, allowing us to understand the spatial structure of the urban system. These patterns arise from interactions between sub-components, but cannot be reduced or simplified – an important property of a complex system.

6.3.Scale-Free Networks: Implications and Cautions

In Section 5.3 we illustrated that the spatial interaction networks formed from bus trip data in Singapore closely resemble that of a scale-free network with a power law degree distribution. Many real-world networks such as a metabolic network and world wide web network are similarly scale-free. Scale-free networks share some of the important characteristics:

a) Absence of an internal scale. In a scale-free network, the characteristics of the network, such as its underlying internal structure and degree distribution, remain unchanged regardless of the network size.
b) Presence of "hubs" in the network. These are nodes with high degrees that are primarily
c) Hierarchical organisation: robust to failure
d) Self-organisation.

The observation of scale-free networks was first reported by Barabasi & Albert in 1999 in a number of real-world networks, such as WWW, actor collaboration network, and power grid network. In the formation mechanism proposed by Barabasi & Albert (1999), as the network grows, new vertices are continuously added and linked to existing vertices in a preferential manner: the likelihood of being linked to an existing vertex that is already well connected is higher than that of a vertex that is less well connected. By extending the explanation of this proposed mechanism, we could understand the formation of the scale-free spatial interaction network as:

- The spatial interaction network grows as more passengers from different locations start to participate in bus trips;
- Choice of trip destination is preferential – passengers are more likely to end their trips at a bus stop that is already a popular destination.

However, we also noted that the assumptions based on which Barabasi and Albert (1999) proposed the formation mechanism do not always apply in this study. For example, Barabasi and Albert (1999) assumed that the network is constantly growing, and new vertices are continuously added. However, the spatial interaction network in a diurnal temporal timeframe shows a distinct lifecycle from birth to death. Hence, Barabasi and Albert (1999)'s proposed mechanism is more suitable for the early period of the diurnal lifecycle, although we observed similar scale-free state as the network gradually contracts and breaks down. It may be worth testing on a longer timeframe to observe the growth and development of the land use and transport systems in conjunction to better explain the formation of scale-free spatial interaction networks.

Although we performed linear regression to acquire an estimation of the gradient, or the power exponent of the long-tailed distribution, the logarithmic degree distribution plot in Figure 5-6(b) shows that the power exponent may not be linear. In fact, we could hypothesise that lower-degree nodes may follow a different gradient from the higher-degree node, or that a quadratic model could provide a better explanation for the gradient. In addition, within the field of statistical physics, there exist active debates on how to prove whether a distribution indeed follows the power law, and such proof requires a strict statistical definition of a power law. Scientists have also provided more relaxed definitions of a scale-free network to accommodate data uncertainty as alternatives, such as log-normal distributions (Broido and Clauset, 2019). The implications of the scale-free state of the spatial interaction networks remain to be rigorously ascertained and the implications further investigated.

6.4. Manifestation of Uncertainty

One prominent form of uncertainty in this study lies with the data sources. The passenger volume data has been anonymised and aggregated into trip counts by hour, day type and OD pair. The dataset does not provide the trip count for every day in the month. Instead, LTA explains that the trip count stands for the expected trip count between an OD pair on a "typical weekday" or a "typical weekend". It is postulated that some methods of aggregation and averaging was applied to compute this value, but the exact method is unknown. In addition, trip counts are aggregated within each hourly timeframe (e.g. from 14:00 to 14:59), and the individual trip timings are not released. Therefore, the uncertainty arises from simplification. As data users, we risk losing the granularity of exact temporal distribution and the precision in temporal analysis result.

Since the exact method of computing the trip counts is not released, it is extremely challenging to reconstruct the exact temporal distribution of trips. In the study, we have chosen to use the same temporal resolution that is given by the dataset (i.e. hourly timeframe) to ensure consistency and avoid introducing further uncertainty. We consider this approach sufficient for the purpose of the study, that is to understand diurnal dynamics of the network evolution, as the hourly resolution is a suitable temporal unit to make sense of the changes in relation to urban activities.

6.5. Limitations and further research

This study presents an initial exploration of the passenger volume data by constructing them into spatial interaction networks. Ideally, such spatial interaction networks should reveal, based on people's travel patterns, how places are functionally dependent. However, as transfers are an integral feature of Singapore's transportation system, the final spatial interaction networks constructed from this dataset will reveal more of the correlation between O/D and interim transfers space, than that between the actual origin and destinations of journeys. There have been attempts by the Singapore Urban Redevelopment Authority to chain trips together to form a complete journey to account for a buffer of transfer time (Chua & Wang, 2018). However, such an approach requires a passenger identifier and the data is not publicly available. As a result, our analysis may have over-estimated the influence of transit nodes that serve primarily as transfer space.

Secondly, we only considered bus data, and as shared previously in the hub-and-spoke model, bus services are supposed to be feeder services that bring people to MRT stations, so again this is not fully representative of the travel patterns. Going further, it will be better to construct multi-layer network incorporating not just bus data but also train data in the analysis.

Lastly, because statistical proofs for power law itself is an active field of research in statistical physics, we are only making general observation and approximate it to power law. More rigorous statistical proofs may be required to demonstrate that the distribution really follows the power law.

7. Conclusion

As a fundamentally interdisciplinary study, this project applied concepts and techniques from network science to acquire a system-level understanding of the urban space. By exploring fundamental network properties in a spatiotemporal framework, we illustrated important findings such as a distinct diurnal lifecycle, the scale-free state at almost all hours, temporally sensitive polycentric form, and evolving community structure. Certainly, the study is not without limitations, but these findings provide preliminary insights into spatial organisation of the urban system of Singapore and its temporal evolution patterns, and outlines key directions to investigate for future research.

One consideration concerning the interdisciplinary research between network science and geospatial studies is the validity of applying concepts from network science and deriving meaningful understanding of geographical space from it. Although we could provide reasonable explanation for most of our findings in the study, there are still gaps of understanding when interpreting the findings in the context of urban studies. For example, the definition of a community based on network topology needs to be meaningfully explained in the context of space and spatial organisation. The same challenge applies to other important network properties such as degree distribution and degree correlation. There are a few other concepts common to both disciplines, such as distance, scale, and neighbourhood/community, that have acquired new meanings in network science. More theoretical research on these network concepts would enrich our understanding on their geospatial implications.

8. References

Barabasi, A.-L., 2016. **Network Science**. Cambridge University Press.

Barabási, A.-L., Albert, R., 1999. **Emergence of Scaling in Random Networks**. *Science* 286, 509–512.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. **Fast unfolding of communities in large networks.** *J. Stat. Mech*. 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Brandes, U., 2001. **A faster algorithm for betweenness centrality*.** *The Journal of Mathematical Sociology*, 25, 163–177. https://doi.org/10.1080/0022250X.2001.9990249.

Brin, S., Page, L., 1998. **The anatomy of a large-scale hypertextual Web search engine.** *Computer Networks and ISDN Systems 30*, 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X.

Broido, A.D., Clauset, A., 2019. **Scale-free networks are rare**. Nat Commun 10, 1017. https://doi.org/10.1038/s41467-019-08746-5.

Centre for Liveable Cities, Land Transport Authority, 2013. **Transport: Overcoming Constraints, Sustaining Mobility**. Cengage Learning Asia Pte Ltd, Singapore.
Chua, A. & Wang, S., 2018, **Journey Estimation with Smart Card Data for Land Use Planning.** 3rd Asia Pacific Conference on Complex Systems Design and Management (CSD&M Asia), December 6–7, 2018, Singapore. Accessed on 6 April 2021 from http://www.2018.csdm-asia.net/IMG/pdf/Journey_Estimation_with_Smartcard_Data_for_Land_Use_Planning.pdf.

Fisher, P., Comber, A., Wadsworth, R., 2006. **Approaches to Uncertainty in Spatial Data,** in: Devillers, R., Jeansoulin, R. (Eds.), Fundamentals of Spatial Data Quality. ISTE, London, UK, pp. 43–59. https://doi.org/10.1002/9780470612156.ch3

Lan, T., Li, Z., Zhang, H., 2019. **Urban Allometric Scaling Beneath Structural Fractality of Road Networks**. *Annals of the American Association of Geographers*, 109, 943–957. https://doi.org/10.1080/24694452.2018.1492898

Land Transport Authority, 2020. **Public Transport Ridership: Average Daily Ridership (2015 – 2019).** Accessed 6 April 2021 from https://www.lta.gov.sg/content/dam/ltagov/who_we_are/statistics_and_publications/statistics/pdf/PT_Ridership_2015_2019.pdf.

Ministry of Transport (n.d.). **Fares and Payment System**. Accessed 28 Mar 2021 from https://www.mot.gov.sg/About-MOT/Land-Transport/Public-Transport/Fares-Payment-Systems/.

Wong, K. Y., 2018. **Singapore's Integrated Land Use Planning Experience.** Accessed 29 Mar 2021 from https://slidetodoc.com/download.php?id=2139529.

Urban Redevelopment Authority (n.d.). **Master Plan**. Accessed 29 Mar 2021 from https://www.ura.gov.sg/Corporate/Planning/Master-Plan.

Zhang, Y., Marshall, S., Cao, M., Manley, E., Chen, H., 2021. **Discovering the evolution of urban structure using smart card data: The case of London.** Cities 112, 103157. https://doi.org/10.1016/j.cities.2021.103157

Zhong, C., Arisona, S.M., Huang, X., Batty, M., Schmitt, G., 2014. **Detecting the dynamics of urban structure through spatial network analysis.** *International Journal of Geographical Information Science* 28, 2178–2199. https://doi.org/10.1080/13658816.2014.914521