

# Identifying spatial super-spreader and spatial super-receiver from human movement networks

Name1 Surname<sup>1</sup>\*, Name2 Surname<sup>1</sup>\*

**1** Singapore University of Technology and Design, Singapore

\*These authors contributed equally to this work.

\* correspondingauthor@sutd.edu.sg

## Abstract

The identification of spatially influential and vulnerable places were crucial to the disease control and prevention, especially for a highly contagious disease, such as the currently pandemic COVID-19. While human movement is the key to the disease diffusion of contagious diseases, the population flow networks is a useful material to understand and uncover the spatially influential and vulnerable places, namely the spatial super-spreaders and spatial super-receivers. But, previous studies mainly focused on non-weighted social network to identify super-spreaders, without considering the flow structure. This study aims to develop a calculation framework for uncovering the spatial super-spreaders and -receivers, with the consideration of local flow intensity and two neighborhood diversities (zone and coreness). We used Singapore public transport flow networks as a case study, in which the flow data were separated into weekday and weekend for comparison. The results showed that most spatial super-spreaders were also super-receivers, with some of exception that a small number of locations were only super-spreaders or super-receivers. Those places were prone to disease diffusion, and they had the ability to spread throughout the country. Thus, we believed that these places required more attentions while establishing disease intervention policies and resource allocations.

## Introduction

Until the preparation of this article, the outbreak of Coronavirus disease 2019 (COVID-19, or nCoV-2019, caused by SARS-CoV-2) is still an ongoing issue with the reported cases surpassed 200,000 [1]. The pathology of the COVID-19 remained an open issue [2, 3], and the development of the vaccines and medicines for treating COVID-19 are still under processing. To the current understanding, COVID-19 is more infectious than the 2003 Severe Acute Respiratory Syndrome (caused by SARS-CoV-1) [4, 5], the main transmission method is through respiratory droplets, and the infected patients would experience a (about) 14 days latent period before starting to have symptoms and become infectious [6, 7]. The asymptomatic latent period of COVID-19 and its highly contagious disease properties had made COVID-19 even more difficult to be control and prevent [5].

The outbreak of COVID-19 started in December 2019 at Wuhan City, Hubei Province of China, which started to spread worldwide in January of 2020, lead to the declaration of a Public Health Emergency of International Concern (PHEIC) by WHO [8]. Until the declaration of PHEIC, 7818 cases were confirmed, in which 82 were

out of China cases [8]. In February, the COVID-19 started to spread internationally, especially to the East and South East Asia, and some European countries who had intensive direct population flow from China. The first wave of international spreading occurred partially because of the end of Chinese New Year holiday and the evacuation of citizens from China, or direct interactions with people who had traveled to Wuhan City, China [9]. The countries in this first wave of outbreak included Thailand, Japan, Singapore, South Korea, France, Germany and United Kingdom [10]. Those imported cases had quickly developed to local transmission in most countries. In March, as the outbreak occurred in Italy, Spain, France, and Germany, the epicenter of COVID-19 moved to Europe [1], which were the second wave of outbreak and international pandemic. The second wave outbreak triggered the lock-down in some of the European countries. The purpose of the country or city level lock-down is to introduce social distancing, reduce the chances of imported cases, and to restrict the human movement to stop the disease spreading spatially within the country.

The identification of super-spreaders is a popular topic in social network studies [11–14]. The 20/80 rule was observed in most disease diffusion studies, indicated that about 20% of the people were responsible for the 80% spread of an infectious disease; namely the super-spreaders [15, 16]. The identification of super-spreaders has great theoretical significance and high potential of practical applications in terms of disease control, hence attract large attention in researchers community and public sectors. Previous studies focused on social networks, in which nodes as people and links as their social interactions, to search for the most influential people by using network analysis metrics, including degree, closeness and betweenness centralities, the k-shell decomposition, or the modification based on these measurements. [12, 14, 17–20]. In some recent studies, researchers indicated that the characteristics of neighboring nodes (the semi-local information or local structure) is efficient in quantifying the nodes' spreading capability [21–23]. On the other hand, some studies pointed out that on the search of multiple (top-k) super-spreaders, there were chances that the identified nodes by the local or semi-local measurements might located at the same network community, which would restrict the spreading events occurred only within the same community. Thus, they developed some methods that considered the community structure while identifying the top-k super-spreaders [19, 24]. In summary, previous studies had shown that two node characteristics are important while quantifying the influential power of a node, including the node's local information—the immediate interactions with its neighboring nodes, and the node's position within the network—its connection to the rest of the network, e.g. if it is connected to all of the communities. To integrate the nodes' local indicators and the nodes' global topological index, Fu et al. proposed a two-step framework to develop the index for identifying the super-spreaders [13].

Although the concept of super-spreaders focused on the person based social interactions, it is not practically applicable for the country-wide or city-wide analysis and policy establishing because of the difficulties on acquiring person-to-person interaction co-presence data and person-based policy making for everyone in the whole country or city. Therefore, previous studies attempted to used spatial network to conceptualize the networked interactions between places [25] for the understanding of spatial disease diffusion [26–29]. People move from places to places at all time. The movements would provide the opportunity of infectious diseases spreading as viruses or pathogens could be transferred between people [30–32]. Previous studies stated that the urban structures could be used to rank the concentration of human activity and population density [33–35]. In order to integrate the human movement, individual interactions and the disease spreading model (e.g. susceptible - exposed - infectious - recovered model, or SEIR model), previous studies applied the metapopulation model to simulate the disease diffusion dynamic process [27, 28, 36]. In summary, spatial networks

could be useful on uncovering the spatial structures behind the disease diffusion networks, and provide insights and decision making supports for country-wide or city-wide. But, most of previous spatial network studies on disease diffusion focused on the space-time development and potential impacts of the disease, and none of which made discussion on the most influential geographical units, namely the spatial super-spreaders.

In addition to the concept of super-spreaders, this study would also describe and discuss the super-receivers. While the concept super-spreaders focuses on the ability to spread a disease, the super-receivers concept emphasizes on the vulnerability of receiving the disease [37–39]. In other words, the aim of the identification of super-receiver is to find the most susceptible node within the network. One similar concept in spatial analysis is the low-high outliers, that is, the location with low density of disease cases which is surrounded by high density locations, thus it is more vulnerable as it had higher possibility to get large cases in the next time period [40]. In network analysis field, the concepts of spreaders and receivers appeared in the HITS algorithm [41] as hubs and authorities, respectively. In HITS algorithm, hub describes the level of being a high influential node and authority represents the popularity of being a destination. The spatial super-receivers represents the susceptible locations in a spatial diffusion network. These locations would be more vulnerable within the network as they are the destination of more people, hence having higher possibility of being visited by infected person.

The spatial super-receivers are vulnerable as they are prone to disease infection, and the spatial super-spreaders are dangerous as they could spread the disease quickly to the rest of a country. If a place is both a super-spreader and super-receiver, it would require more attention as it would be easily get infected and quickly spread to other places. Identifying these places would be important and useful in the fight with infectious diseases. Therefore, the aim of this study is to identify the spatial super-spreader and spatial super-receiver in an island country's spatial flow network. A spatial super-spreader is a location where a lot of people are originated from, and those people are moving to a high variety of places; a spatial super-receiver is the destination of a large number of commuters, who come from different places. In this study, we developed two indexes, namely the spreader index (*SI*) and receiver index (*RI*) to search for the spatial super-spreaders (SS) and spatial super-receivers (SR), which considered two key components: the local strength of flows and the diversity of its neighborhood [13]. The local strength of flows for a location is the number of people leaving from or coming to the location, i.e. respectively the weighted out-degree and weighted in-degree. The neighborhood diversities would be captured by two type concepts: one of which is the diversity of zones [34, 42], i.e. are the people come from different parts of the country; another is the diversity of coreness [43, 44], i.e. are the people come from different types of the country in terms of core or peripheral areas. We applied the analysis framework to the Singapore public transport flow network, and identify the spatial super-spreaders and spatial super-receivers using the spreader and receiver indexes. The population flow pattern may be different for weekday and weekend. Thus, the flow data were separated into two parts, weekday and weekend, for the discussion on the differences of super-spreaders and super receivers during weekdays and weekends.

## Materials and methods

The section contain three parts: (a) brief description of the study area, (b) the flow data, and (c) the four steps calculation of the spreader-index and receiver index.

## Study area

This study focused on the public transportation flow network in Singapore. Singapore is an island country located at South-East Asia with a total area of about  $724.2 \text{ km}^2$ . As of 2019, the total population of Singapore is about 5.703 million people (the population density is about 7875.68 per  $\text{km}^2$ ), in which 70.6% are residents (citizen and permanent resident) and 29.4% are non-residents (foreigners with long term pass). According to the General Household Survey 2015, about 62.7% students and 64.1% working person relies on bus or rail transport services to travel to schools or work places, making public transport the major mode of commuting in Singapore. As a result, the density of people in the public transport during the morning peak and evening peak are high, and the distance between people at the stations or vehicles are close. The short social distance raised a lot of concern in the discussion of disease diffusion because it would provide more opportunity for viruses or pathogens to spread between people.

This study used the subzone level spatial boundaries (from Master Plan 2014) as the analysis unit. The residential population density (from General Household Survey 2015) were shown in Fig 1. There were five regions (Central, West, North, North East, and East), 55 planning areas, and 323 subzones. Some of the subzones contain no residential population (white areas), which includes airports and airbases (e.g. Changi Airport at the East Region) and industrial parks or ports (e.g. Jurong Island and Bukom at the south of the West Region, and Simpang North and South at the North Region). Although these places contain zero residential population, they could be the work places (destinations) of a lot of commuters. The darker color areas indicate the home for a large number of people; in other words, a large number of journeys started from and ended at these locations.

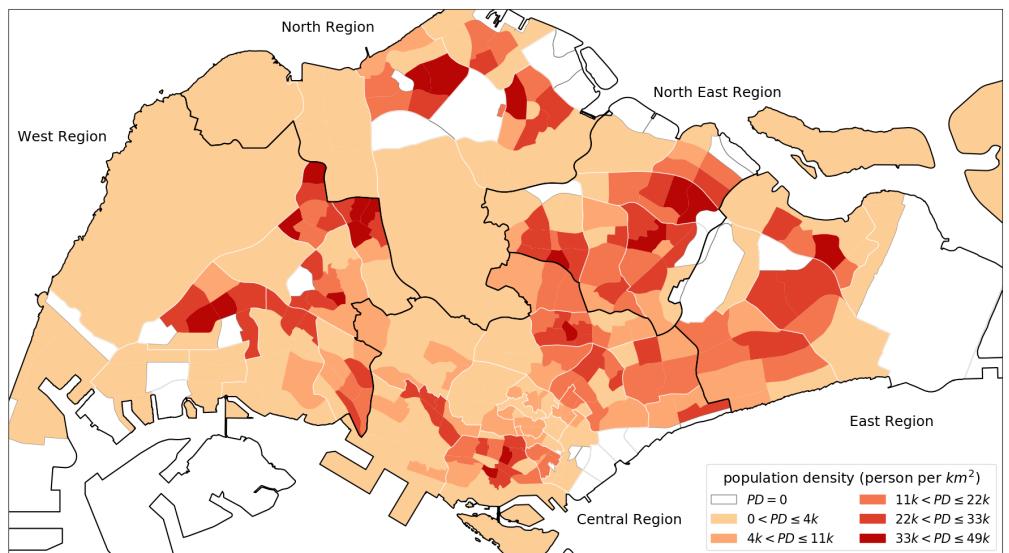


Fig 1. The subzone residential population density map of Singapore.

## Weekday and weekend flow network

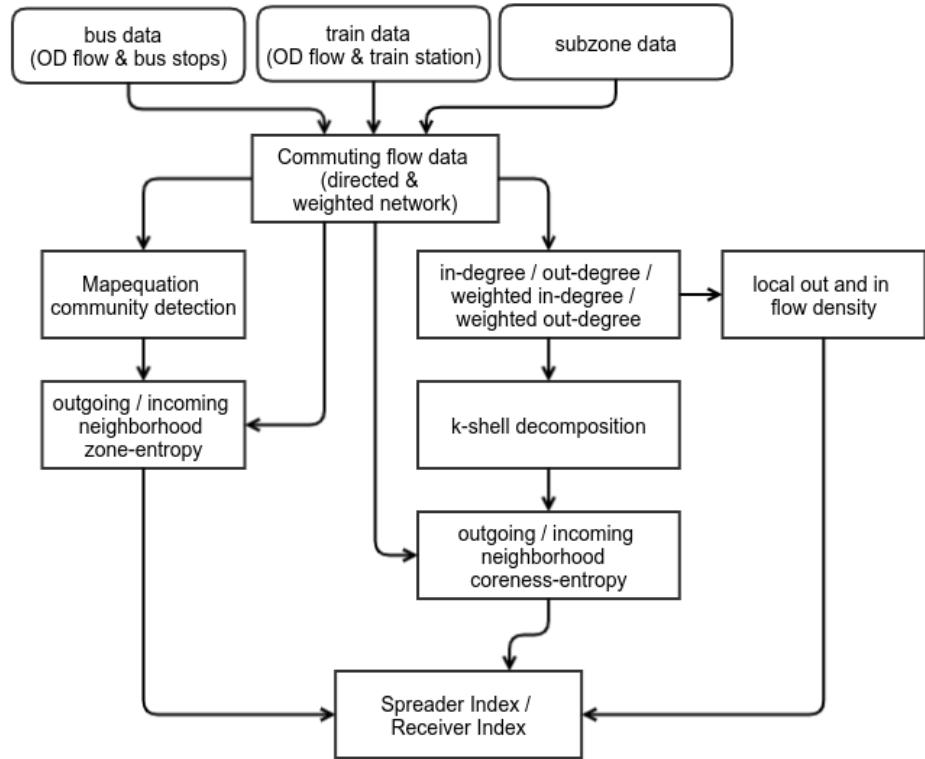
We used the origin-destination (OD) ridership data of bus and train to create public transport flow networks. The OD ridership data were collected from the Singapore Land Transport Authority (LTA) through API calls. In this study, we used the ridership in January 2020. The OD ridership data contained the hourly passenger flows between each pair of bus stops or train stations (including mass rapid transit and light

rail transit). The data were aggregated into weekday (a total of 23 days in January 146  
2020) or weekend (8 days). 147

As the raw data recorded the flow between the OD pair of bus stops or train 148  
stations, we spatially aggregated the data into subzones to subzones flows, according 149  
to the bus stops or train stations locations. A total of 303 subzones contained at least one 150  
bus stop or one train station. These subzones were used as the nodes (303 nodes) in the 151  
flow networks, with the flows between them as the weight of directed edges. A total of 152  
30331 edges were found, of which 30043 were inter-subzones edges, and 288 were 153  
within-subzone flows (self-loops). The within subzone flows were ignored in this study. 154

## Calculation framework 155

The calculation flow of the spatial spreader and receiver indexes is shown in Fig 2. The 156  
first part is to aggregate the bus and train OD flow data to subzones as aforementioned. 157  
Then, we got the main data for the calculation, i.e. two weighted and directed networks: 158  
weekday and weekend flow network. These networks were used to calculate three 159  
network characteristic measurements, including degree centralities (step 1), community 160  
detection (step 2), and k-shell decomposition (step 3), which were described in the 161  
following subsections. The degree centralities were used as the local out and in flow 162  
intensity, whereas the community detection and k-shell decomposition results were used 163  
to calculate the neighborhood diversities, including zone-entropy and coreness-entropy. 164  
Finally, the three network characteristics were used to calculate the spreader index and 165  
receiver index (step 4). 166



**Fig 2.** The calculation flow chart of the spreader and receiver index.

## Step 1: Degree centralities

The degree centralities in this study includes the non-weighted and weighted for both in and out degrees. The non-weighted and weighted versions of degree centralities represent different concepts in network characteristics. The non-weighted in-degree and out-degree are the number of link that is pointed to and from a subzone, respectively. These non-weighted degree centralities measure the number of relationships a subzone has. The weighted in-degree and out-degree are the summation of incoming flows and the summation of outgoing flows of a subzone, respectively. This weighted version of degree centralities indicate the total strength of a node in terms of gathering flows or spreading flows, but it do not differentiate the number of links.

In this study, the weighted degree centralities were used to represent the local intensity of nodes for the calculation of spreader and receiver indexes. The weighted degree centralities were scaled to the range between 0 and 1 for the calculation. On the other hand, both non-weighted and weighted degree centralities were used in the weighted k-shell decomposition analysis in step 3.

## Step 2: Zone-entropy

This study used community detection method [42] to identify the zones from the flow network, instead of using the political spatial boundaries (the boundaries of planning areas and regions from Master Plan 2014) that were design for governance purpose. The communities from flow network capture the strength and direction of flows, which showed the activity space of people that derived from the commuting behaviors [34]. As the community distribution was identified from weekday and weekend networks, the distribution should be different between weekday and weekend.

MapEquation is used to identify the communities in the flow networks [42]. MapEquation is an algorithm that consider the direction and weight of edges to identify the strongly connected nodes in a directed and weighted network. Different from modularity-based community detection methods, MapEquation's calculation concept emphasize the strength of flows in community, i.e. more flows were moving within a community than between communities (flows cycling within communities).

MapEquation captures the effect of direction while ensuring large amount of flows are kept within community. The MapEquation communities are used as the zones that contain a strong population flows cycle, and are used to calculate the zone-entropy.

First, we ran the MapEquation on the two networks, and identify the zone (community) each subzone (node) is belonged to. Then, for each subzone, the incoming/outgoing neighbors' zones were retrieved from the result, together with the weights of incoming/outgoing edges. The neighbors' zone information and flow weights were used to calculate the normalized entropy ( $H_{Neigh}^{Zone}(i)$ ) using Eq (1)-(3). In which, the entropy is normalized using the total number of zones in the network, so it could be compare between each node. The zone-entropy value range is between 0 and 1.

$$H_{Neigh}^{Zone}(i) = \frac{-\sum_{Z \in Zone(OutNeigh)} P_i(Z) \ln P_i(Z)}{\ln |Zone(All)|} \quad (1)$$

$$Neigh = \{OutNeigh, InNeigh\} \quad (2)$$

$$P_i(Z) = \begin{cases} \frac{\sum_{j \in Z \cap Neigh(i)} w(i,j)}{\sum_{k \in Neigh(i)} w(i,k)}, & \text{if } Neigh = OutNeigh \\ \frac{\sum_{j \in Z \cap Neigh(i)} w(j,i)}{\sum_{k \in Neigh(i)} w(k,i)}, & \text{if } Neigh = InNeigh \end{cases} \quad (3)$$

### Step 3: Coreness-entropy

K-shell decomposition is a method to label the coreness (k-shell levels) of nodes in a network based on the connectivity structure [12]. Because the edges of the flow networks were weighted, we used the weighted k-shell decomposition [43], which was an extended version that consider both the number of links (degree) and the weights of links while labeling coreness. The coreness of a location indicates the position of the location in the range from periphery (low k-shell levels) to core (high k-shell levels). In a population flow network, the core locations indicate the common origins or destinations for a large number of passengers.

In this study, we first ran the weighed K-shell decomposition using the non-weighted and weighted in/out degree (from step 1) to calculate the in/out-k-shell level for each subzone. Then, the k-shell levels were grouped into core (in/out-core) or periphery (in/out-non-core) using the median value as a break. Finally, for each node, its incoming/outgoing neighbors' core/non-core information were integrated with the flow weights to calculate the coreness-entropy ( $H_{Neigh}^{Core}(i)$ ) using Eq (4)-(6). As shown in the denominator of Eq (4), the entropy is normalized using the total number of coreness levels (which is 2, i.e. core or non-core), so the result could be compare between nodes. The coreness-entropy value range is between 0 and 1 after the normalization.

$$H_{Neigh}^{Core}(i) = \frac{-\sum_{C \in Core(Neigh)} P_i(C) \ln P_i(C)}{\ln |Core(All)|} \quad (4)$$

$$Neigh = \{OutNeigh, InNeigh\} \quad (5)$$

$$P_i(C) = \begin{cases} \frac{\sum_{j \in C \cap Neigh(i)} w(i,j)}{\sum_{k \in Neigh(i)} w(i,k)}, & \text{if } Neigh = OutNeigh \\ \frac{\sum_{j \in C \cap Neigh(i)} w(j,i)}{\sum_{k \in Neigh(i)} w(k,i)}, & \text{if } Neigh = InNeigh \end{cases} \quad (6)$$

### Step 4: Spatial spreader & receiver index

The spatial spreader index ( $SI$ ) and receiver index ( $RI$ ) are modified from the framework in [13]. The  $SI$  and  $RI$  calculation are the cube root of the multiplication of the three of the aforementioned network measurements. The  $SI$  (Eq (7)) is calculated as the cube root of the multiplication of the local normalized weighted out-degree ( $NWOutDegree(i)$ ), the zone-entropy of outgoing neighbors ( $H_{OutNeigh}^{Zone}(i)$ ), and the out-coreness-entropy of the outgoing neighbors ( $H_{OutNeigh}^{Core}(i)$ ). As a result, if a node's  $SI$  is high, it has a high volume of outgoing flows (high local intensity), half of the flows were going to core area and half to non-core area, and these flows are equally divided into different zones (high out-neighbors' zone-entropy). In other words, a high  $SI$  subzone has a large number of travelers originated from there, and they are going to both core and periphery places, which are located in varying zones. Therefore, this kind of origins would have stronger ability to spread disease within a short time period. The flow intensity and diversity measurements are all in the range between zero and one, thus after the multiplication and cube root, the result would also in between zero and one.

$$SI(i) = \sqrt[3]{NWOutDegree(i) \times H_{OutNeigh}^{Zone}(i) \times H_{OutNeigh}^{Core}(i)} \quad (7)$$

The  $RI$  (Eq (8)) is calculated as the cube root of the multiplication of the local normalized weighted in-degree ( $NWIInDegree(i)$ ), the zone-entropy of incoming neighbors ( $H_{InNeigh}^{Zone}(i)$ ), and the in-coreness-entropy of the incoming neighbors ( $H_{InNeigh}^{Core}(i)$ ). A high  $RI$  indicates that the subzone has large incoming flows, half of

the flows are coming from core area and half from non-core area, and these flows were  
equally coming from different zones. In other words, this subzone is a destination for a  
large number of travelers, they are coming from various zones and their origin of  
movements contain both core and periphery areas. Therefore, a high *RI* subzone is  
more vulnerable and sensitive in terms of easily get infected. The *RI* is also in between  
zero and one.

$$RI(i) = \sqrt[3]{NWI_{InDegree}(i) \times H_{InNeigh}^{Zone}(i) \times H_{InNeigh}^{Core}(i)} \quad (8)$$

## Results

The results were showed in the following five parts: degree centralities, community  
detection, coreness, spreader and receiver indexes, and the super-spreaders and  
super-receivers.

### Part 1: The local intensity of flow

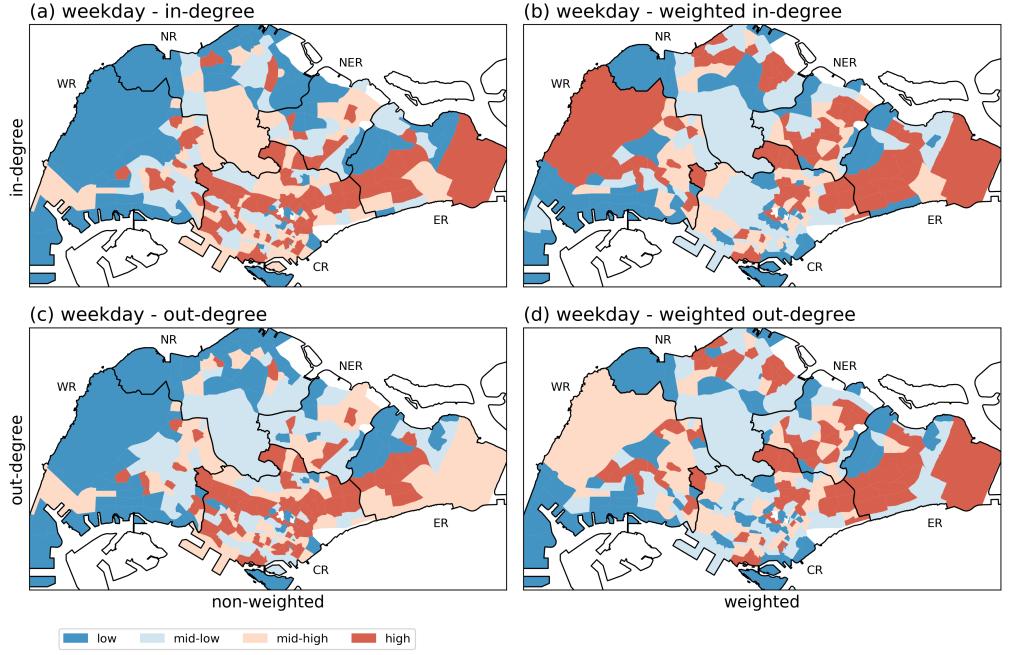
The spatial distribution of the non-weighted/weighted in-degree and out-degree for  
weekday are shown in Fig 3. The distribution patterns for non-weighted and weighted  
in-degree (first row) are similar to their out-degree counterpart (second row). For  
non-weighted degree measurements, the high in- and out-degree subzones mainly  
concentrated at the East and Central Regions, whereas the West, North, and North  
East had more lower degrees subzones. For weighted degree measurements, the East  
Region had higher degrees subzones; the number of high degrees subzones dropped at  
Central Region; North, North East, and West Regions had more high degrees subzones.

The distribution of the non-weighted measurements for weekend were same as the  
results of weekday. In Fig 4, we presented the differences of weighted in- and out-degree  
between weekday and weekend. Most subzones were in the lightest green or blue colors,  
indicated that their degree measurements were only slightly larger than each other  
(lower than 1.3 times). These subzones had similar number of people using the public  
transport for weekday and weekend. Only a few subzones were in dark colors. These  
subzones indicated the usage of public transport at these locations changed between  
weekday and weekend; the changes of usage for weekday are twice larger than weekend's  
(dark purple), or the other way around (dark green).

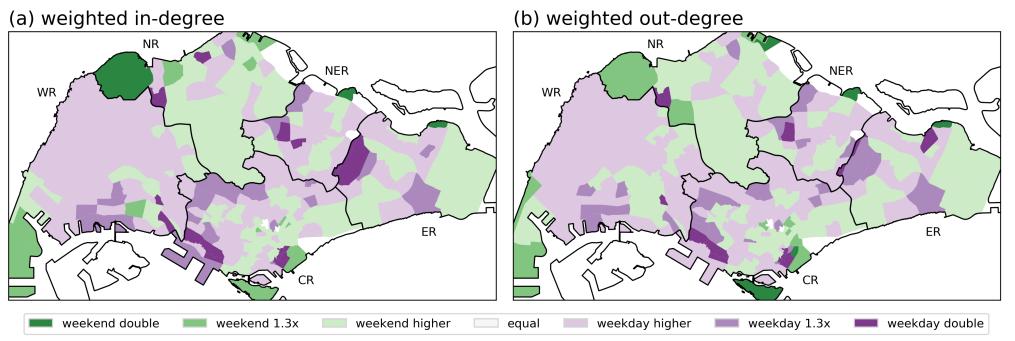
### Part 2: Community detection

Fig 5 showed the spatial distribution of the communities. There are 10 communities in  
the weekday flow network, and 11 communities in weekend flow network. Most  
communities were spatially continuous, but some exception existed in both weekday and  
weekend communities. For example, the no. 4 and no. 8 communities in weekday result,  
and no. 2 and no. 7 communities in weekend result. The spatially continuous pattern  
indicates that the inter-subzones flows were stronger between closer subzones. The  
spatially separated situation indicates that a strong flows of people were moving  
between the two parts of community. This might happen when one of the part have a  
public transportation interchange that attract a larger flow of people. The weekend  
communities were smaller and scattered while compare to the weekday result, which is  
spatially larger in overall.

From the community detection results, the zones for weekday and weekend were  
slightly different with some similarity. On both results, the east area (no. 1 in weekday  
and no. 2 in weekend) were similar, which area also resemble the East Region boundary.

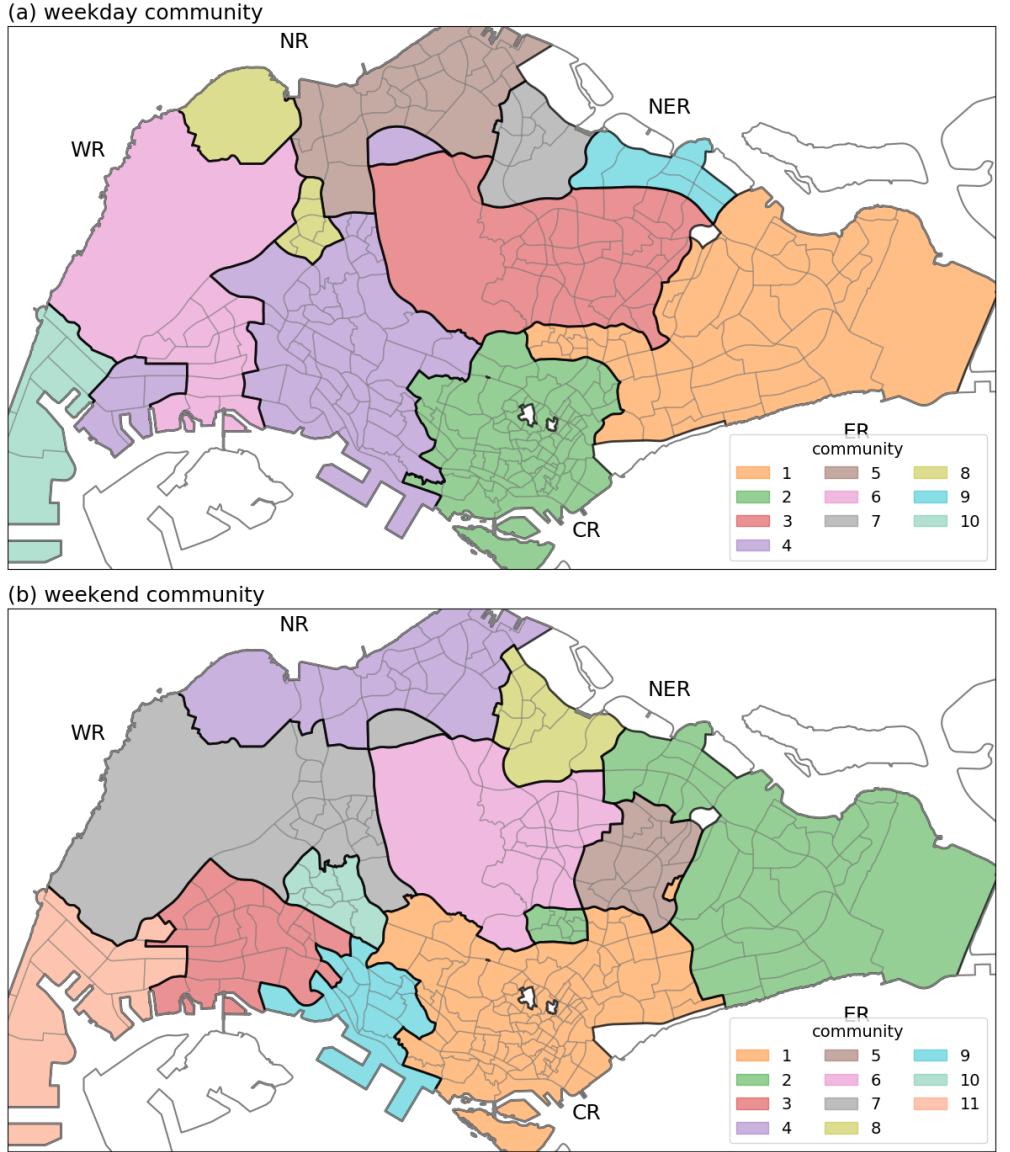


**Fig 3. The spatial distribution of the degree centralities for weekday dataset.** First column (a, c) showed the distribution for non-weighted measurements and second column (b, d) showed the distribution for weighted measurements of degree. The first row (a, b) showed the in-degree, second row (c, d) showed the out-degree. The townships were separated into four groups using 25%, 50% and 75% percentile as breaks.



**Fig 4. The differences of weighted in- and out-degree between weekday and weekend.** Subzones in green indicates weekend had higher degree, whereas subzones in purple indicates weekday had higher degree. The color range from light to dark indicates the scale of higher from one to another.

The communities at the Central Region (no. 2 in weekday and no. 1 in weekend) were smaller than the Central Region boundary. The West, North, and North East Regions were separated into multiple communities. The community detection results showed that the boundaries of human activity were changed between weekday and weekend. For example, the boundaries of the central area were smaller on weekday and larger on weekend; one of the communities at North East Region on weekday were merged to the large community at East Region on weekend. In summary, the human movement boundaries were not fixed to a static shape, and it did not always follow the shape of the political boundaries.



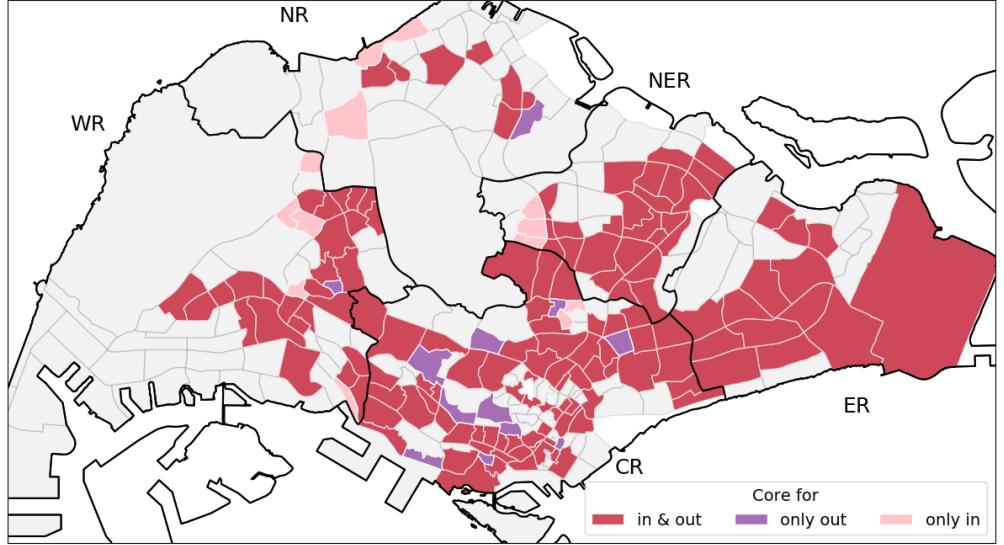
**Fig 5. The distributions of resulting zones from MapEquation.** The detected communities for (a) weekday flow data and (b) weekend flow data. The colors indicate the communities. The white color subzones were ignored in this study because no operating bus stops or train stations were found in the data.

### Part 3: Coreness

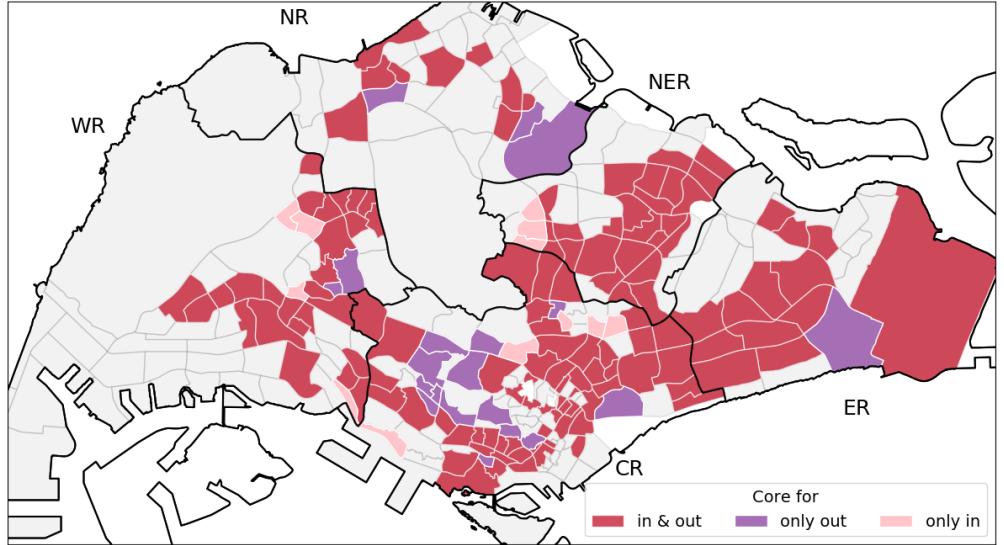
The spatial distribution of core area are shown in Fig 6. The calculation of coreness were separated into two parts for each network, one of which used in-degree and weighted in-degree, and another used out-degree and weighted out-degree. So two sets of coreness results (outgoing core area and incoming core area) were computed for each network. Some areas were identified as core in both incoming and outgoing directions (subzones in red), and some were core for either incoming (in pink) or outgoing (in purple) but not both. Most of the cores were overlap between incoming and outgoing directions. These areas were also overlapped with the high residential population

density area, indicating that places where the people live would always have high incoming and outgoing flows thus becoming the core area of commuting. In weekday result, the out only core area were concentrated at the Central Region, which was the main central business area of Singapore; whereas the in only core area were scattered at North, West and North East Regions. While the Central Region contained the most of out only core in weekend result, the spatial distribution was relatively scattered through out all regions. The in only core area were also scattered as weekday result.

(a) weekday coreness



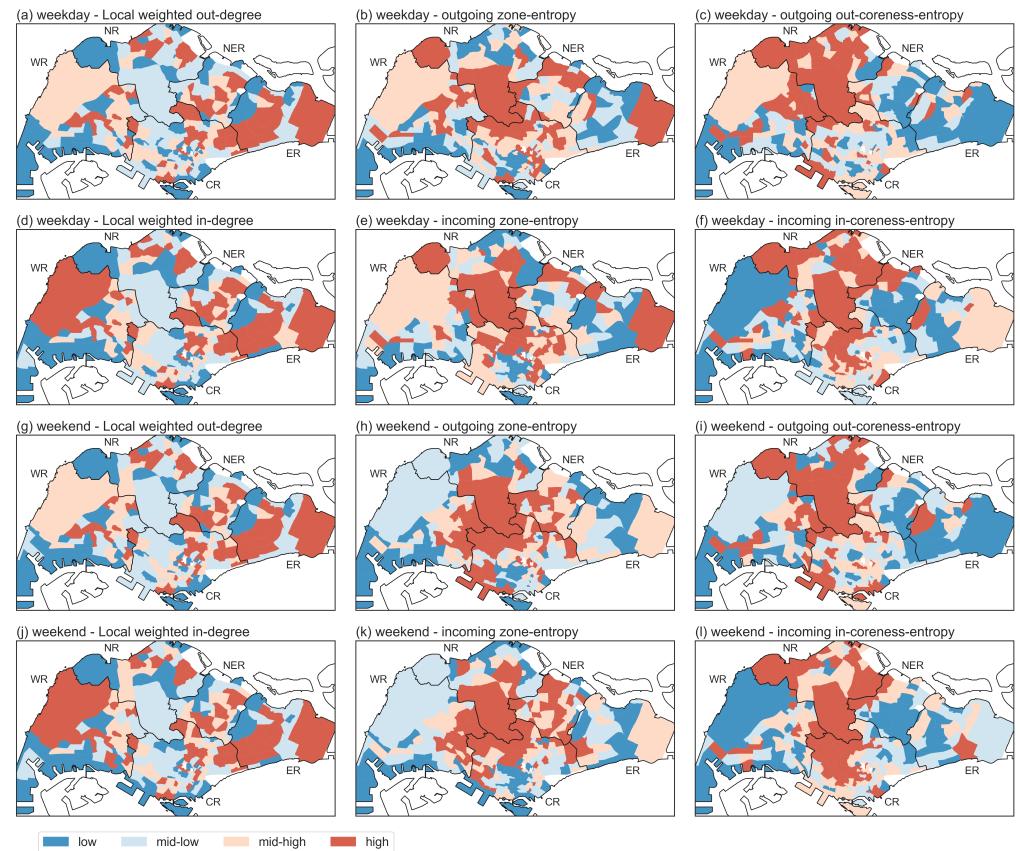
(b) weekend coreness



**Fig 6. The distributions of resulting core/non-core from weighted k-shell decomposition.** The coreness of for (a) weekday flow data and (b) weekend flow data. Red color area showed the subzones which were identified as core area for both incoming and outgoing direction, purple color area showed the outgoing core subzones, and pink color showed the incoming core subzones.

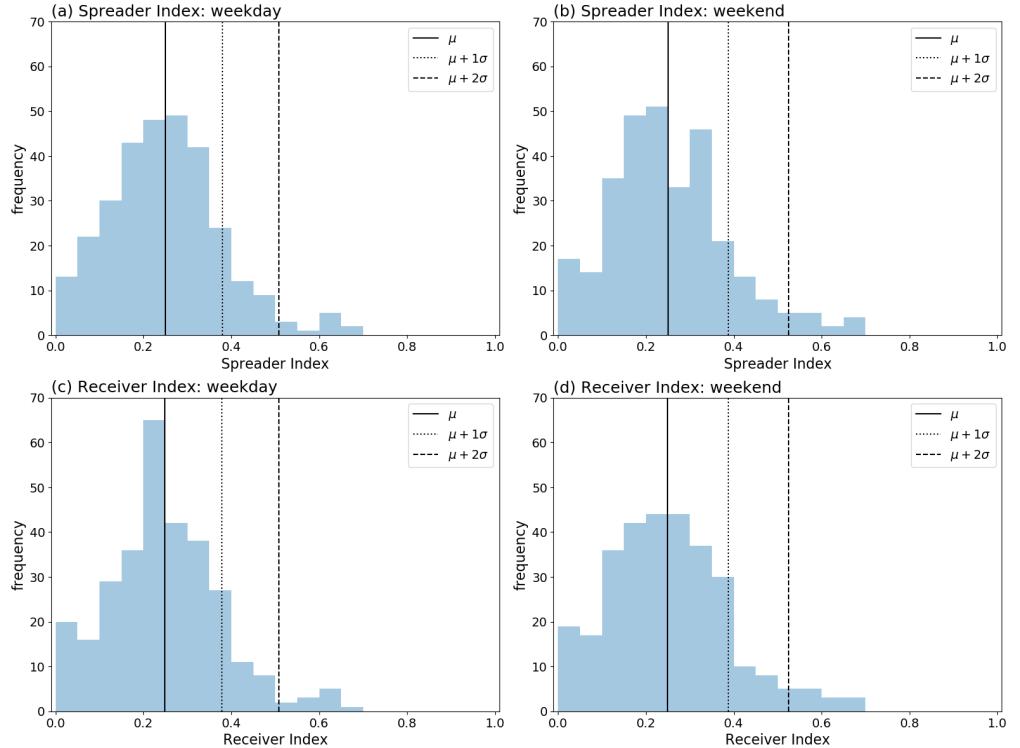
## Part 4: Spreader and receiver indexes

The calculation of spreader and receiver indexes required the local normalized in-degree and out-degree centralities, incoming and outgoing neighborhood zone-entropy and coreness-entropy. These six node (subzone) characteristics were range between zero and one. Fig 7 shows the local out- and in-degree (first column), the outgoing and incoming neighborhood zone-entropy (second column) and coreness entropy (third column) of the weekday (first two rows) and weekend (the last two rows) flow network. The data in each row (three variables) were used to calculate either spreader index or receiver index. The spatial distribution showed that variables in the three columns were different. The high local weighted out- and in-degree were mostly concentrated at the East, West, and North East Regions; the high neighborhood zone entropy were mainly located at the North and Central Regions; most of the highest neighborhood coreness entropy subzones were at the North Region. In other words, most of the subzones contained high values of either one or two of the three variables; only those subzones with high values of the three variables were the super spreader or super receiver.



**Fig 7. The spatial distribution of the 6 variables of weekday (a-f) and weekend (g-l).** First two rows (a-f) were results for weekday and the last two rows (g-l) were for weekend population flow. a and g showed the local weighted out-degree; b and h showed the zone-entropy of the outgoing neighbors (outgoing); c and i showed the outgoing-coreness-entropy of the outgoing neighbors. These three variable were used to calculate the spreader index. The d and j showed the local weighted in-degree; e and k showed the zone-entropy of the incoming neighbors (incoming); f and l showed the incoming-coreness-entropy of the incoming neighbors. These three variables were used to calculate the weekend receiver index.

The distribution for the spreader index ( $SI$ ) and receiver index ( $RI$ ) of each subzone in weekday and weekend were shown in Fig 8. All the four distributions showed bell shape pattern indicating normal distribution, with a mean at about 0.248 to 0.252 (solid lines),  $\mu + 1\sigma$  about 0.378 to 0.388 (dotted lines), and  $\mu + 2\sigma$  about 0.507 to 0.525 (dashed lines). The subzones that lay above the  $\mu + 2\sigma$  threshold were identified as super-spreader or super-receiver, which number were about: (a) 10 weekday super-spreader, (b) 13 weekend super-spreader, (c) 11 weekday super-receiver, and (d) 13 weekend super-receiver. The subzones that lay between  $\mu + 1\sigma$  and  $\mu + 2\sigma$  were categorized as secondary-spreader or secondary-receiver.



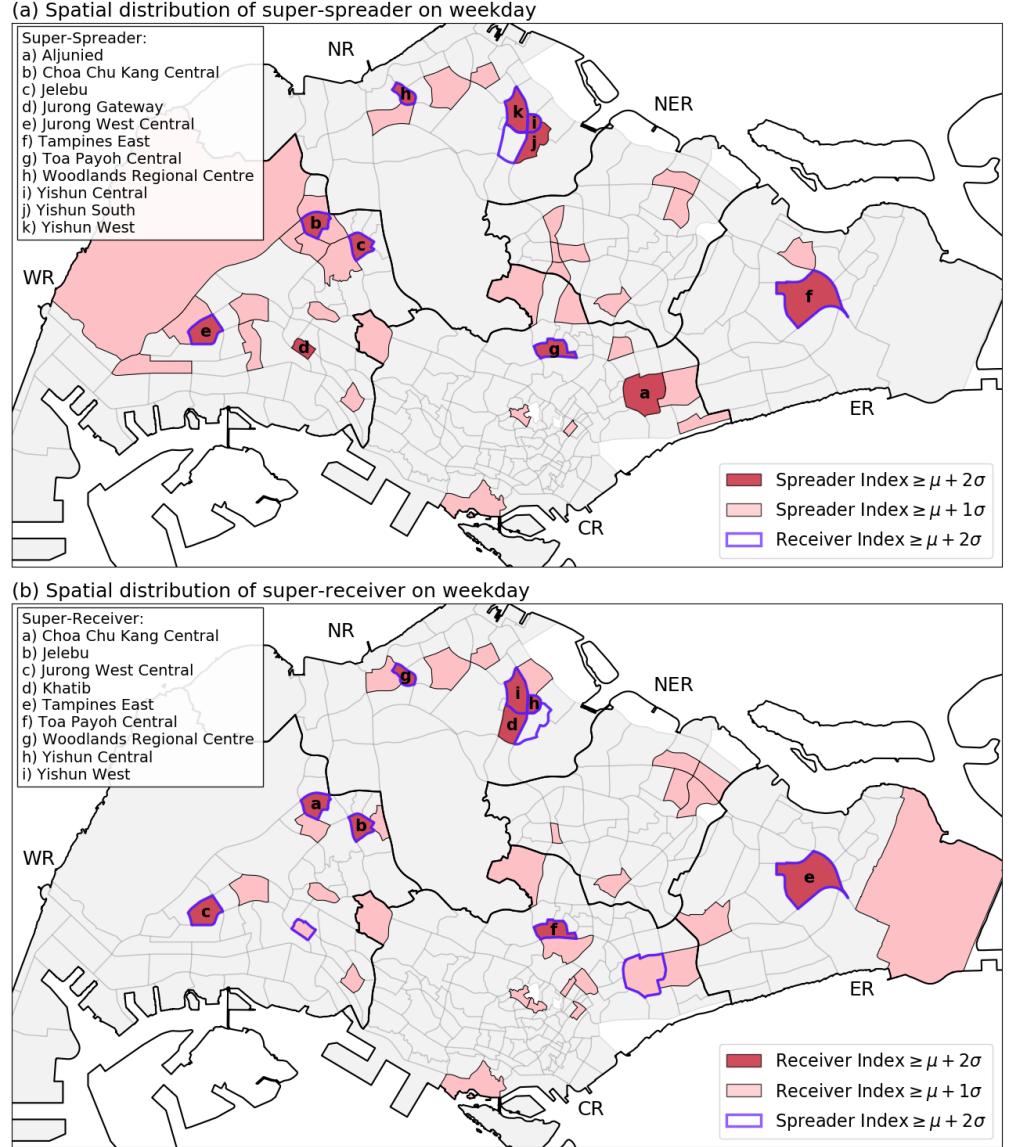
**Fig 8. The frequency distribution of the spreader index and receiver index.** The frequency distribution of the spreader index (a, b) on the first row, and receiver index (c, d) on the second row, for the two datasets: first column (a, c) for weekday, and second column (b, d) for weekend. The vertical solid lines indicated the mean of the distributions, and the vertical dashed lines showed the two times of standard deviation larger than the mean of the distributions. The subzones lie outside the dashed lines are the subzones with the highest spreader or receiver indexes, which were identified as the super-spreaders and super-receivers.

## Part 5: Super-spreader and super-receiver

The spatial distribution of the super-spreader (SS) and super-receiver (SR) were shown in Fig 9 for weekday, and in Fig 10 for weekend. In the weekday result (Fig 9), 11 subzones were identified as super-spreader (red colored with labels in Fig 9a,  $SI \geq \mu + 2\sigma$ ); nine subzones were identified as super-receiver (red colored with labels in Fig 9b,  $RI \geq \mu + 2\sigma$ ); eight of them were overlapped as both super-spreader and super-receiver (red colored subzones with purple border). This indicated that most of the subzones with the highest spreader index would also had the highest receiver index,

and vice versa. In Fig 9a, three subzones ((a) Aljunied, (d) Jurong Gateway, and (j) Yishun South, the red colored area without purple border) were identified as super-spreader, but not as super-receiver; two of them (a and d) were secondary receiver ( $\mu + 1\sigma \leq RI < \mu + 2\sigma$ ); Yishun South's  $RI$  was less than  $\mu + 1\sigma$ . In Fig 9b, one subzone ((d) Khatib) were identified as super-receiver but which spreader index was less than  $\mu + 1\sigma$ .

344  
345  
346  
347  
348  
349



**Fig 9.** The spatial distribution of (a) spreader index ( $SI$ ), and (b) receiver index ( $RI$ ) for weekday. The subzones with purple border in (a) and (b) respectively indicate the super-receiver ( $RI \geq \mu + 2\sigma$ ) and super-spreader ( $SI \geq \mu + 2\sigma$ ).

The weekend distribution result showed slightly different patterns. There were 12 subzones identified as super spreader on weekend. Nine of them were identified as super spreader on weekdays (c, e-l in Fig 10a), and the other three were secondary spreader (a, b, and d in Fig 10a); none of which were less than  $\mu + 1\sigma$  in the previous figure. Some

350  
351  
352  
353

of the weekday super spreader (a and j in Fig 9a) became secondary spreader on weekend. Similarly, all weekend super receiver on weekend were either super or secondary receiver on weekdays, and vice versa. A total of 12 super receivers were found on weekend, 8 of which were overlap with weekday super receiver result (c-d and f-l in Fig 10b), the other four of which were promoted from weekday secondary receiver (a, b, c, and e in Fig 10b); one of the weekday super receiver became secondary receiver (a in Fig 9b). This situations indicated that the spreader indexes and receiver indexes were not dramatically change between weekday and weekend. The differences between spreader and receiver indexes were four subzones: Boulevard and Khatib (a and g in Figure 9b) were super receivers, but Khatib had a spreader index less than  $\mu + 1\sigma$ , whereas Boulevard was a secondary spreader; Ang Mo Kio Town Centre and Choa Chu Kang Central (a and c in Fig 10a) were super spreaders, but they were only secondary spreader on weekend.

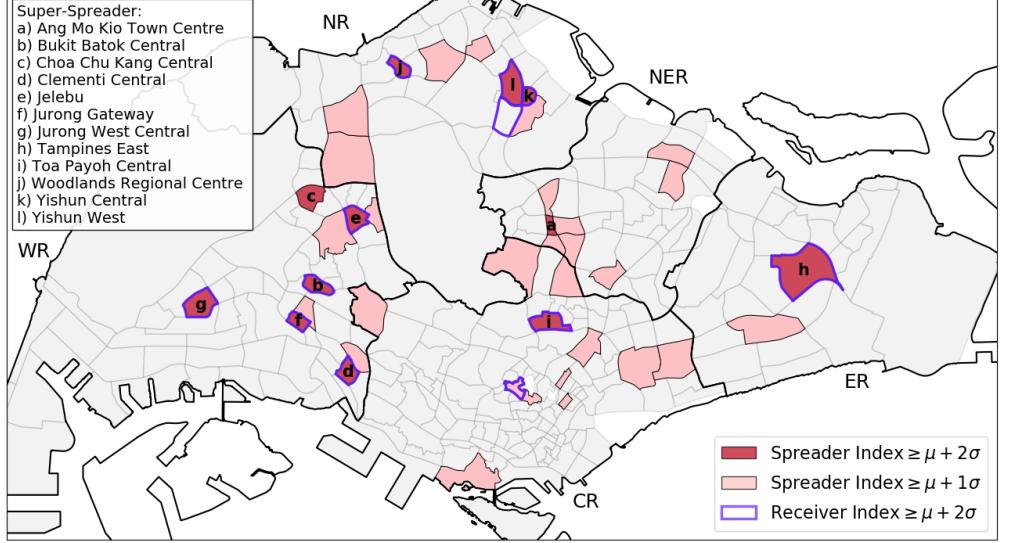
Seven subzones (Jelebu, Jurong West Central, Tampines East, Toa Payoh Central, Woodlands Regional Centre, Yishun Central, and Yishun West) were identified as both SS and SR (in red) in both weekday and weekend. During weekday, most of the identified SS or SR area were the regional core area that contained a higher density of human activity. Tampines East (d) and Aljunied (i) had a relatively large area. Although their residential population density were not among the highest population area in the country, they contained a main bus interchanges, multiple MRT stations, and several shopping malls. These places were the transport hubs, e.g. Tampines East contained an interchange of two MRT lines, whereas Aljunied contained several MRT stations and near to a MRT interchange. Tampines East remained in SS and SR in weekend, while Aljunied became less important of the role as spreaders and receivers. Choa Chu Kang Central (a), Jelebu (b), Jurong West Central (c), and Woodlands Regional Centre (f) were SS and SR on weekday and weekend. These places had a smaller area but high residential population density. Jurong Gateway (j on weekday and d for weekend) was a small subzone with a low amount of population, but it was a SS on weekday and both SS and SR on weekend. This may because it contained the Jurong East MRT station, a large bus interchange, and several shopping malls, which made it a core of daily activity with large intensity of population interactions.

One counter intuitive result was shown from Figure 8 and 9, which is the central business districts (CBD) contain less SS and SR as expected. The CBD of Singapore was located at the central part of the Central Region. Those areas had a high density of human activity. As shown in Figure 7, most of the subzones in CBD had either low weighted degree, neighborhood zone-entropy, or neighborhood coreness-entropy. The low weighted degree probably cause by the area were small, which limited the catchment of incoming or outgoing flows; the low zone-entropy was caused by the majority of the people were circulating within the CBD, which subzones were mainly in the same zone (zone 2 in weekday and zone 1 in weekend, Figure 4); the low coreness entropy was caused by the majority of the flows were moving to or from the subzones in CBD, which were mainly composed by the core area (Figure 5). This result indicates that the CBD work places were less influential in terms of quickly spreading the disease to the rest of the country, but they would be quickly spread inside CBD area as a consequence of the strong flows within the CBD. The more influential area were the regional transport hubs, which were near to the residential areas.

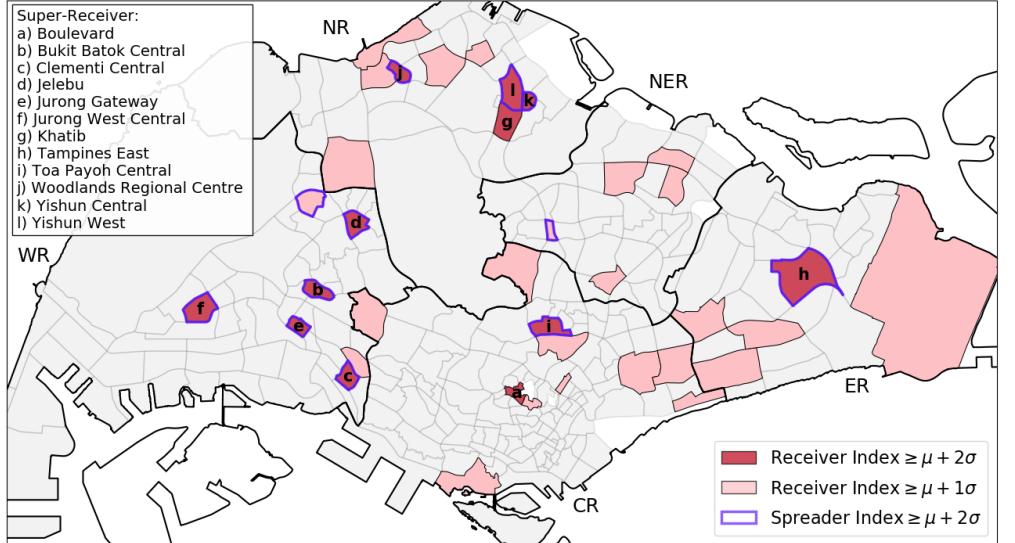
## Discussion

The concept of super-spreader was studied in social network field to identify the most influential person or node within a social network. This person could be an opinion leader within a group of people (e.g. [45]). On the other hand, super-spreader was also

(a) Spatial distribution of super-spreader on weekend



(b) Spatial distribution of super-receiver on weekend



**Fig 10.** The spatial distribution of (a) spreader index ( $SI$ ), and (b) receiver index ( $RI$ ) for weekend. The subzones with purple border in (a) and (b) respectively indicate the super-receiver ( $RI \geq \mu + 2\sigma$ ) and super-spreader ( $SI \geq \mu + 2\sigma$ ).

used in epidemiology study to identify the potential super-spreader of diseases within a social group (e.g. [46]). While previous studies focused on the identification of super-spreaders from a social network, i.e. nodes were person, and edges represented the existence of interactions between a pair of people (binary edge), this study focused on the population flow spatial network with nodes represent locations and weighted/directed edges represent the population flows. This study intended to apply the super-spreader concept to spatial interaction networks, to identify the spatial super-spreaders, which is a set of location that had the most influential power in terms of spreading diseases. The concept and calculation method were also reversed to find the most vulnerable places, namely spatial super-receiver.

The results showed that most of the SS are also SR. This is reasonable because in a daily population flow network, the number of people who are leaving from a place is usually proportional to the number of people who are going to the place, i.e. the larger the outgoing flow intensity, the larger the incoming flow intensity. Based on the results, the places with large amount of flows had higher potential to be both SS and SR, and this was captured by including the weighted in-degree or out-degree in the calculation. This result agreed with previous studies which were based on k-shell decomposition method, that the core nodes of a social group were usually the most influential nodes [12, 18].

Beside the local incoming and outgoing flow intensities, this study also considered two neighborhood diversities of network characteristics: the zone-entropy and coreness entropy. The diversity of neighborhood were specifically important while identifying multiple super-spreaders from a network [13, 24]. The zone-entropy was used to measure if the outgoing flows were moving to more zones within the country. If the outgoing flows from a place are moving to places in only one zone, this place can only affect one of the zone among all the zones in the country, thus its influential power is relatively weak; if the flows were moving to places of all zones across the country, its influential power is stronger. On the other hand, coreness entropy captures the diversity of flows to or from core or periphery area. If the flows were all moving to one of the periphery or core, its influential power was limited to the type of area; if the flows were divided to both core and periphery areas, this indicated that whenever an outbreak happens at this place, it could quickly affect both core and periphery areas. These two diversity measurements were used in the calculation framework for differentiating the places with high density of flows into strong and weak influential places.

This study provided a list of subzones which had stronger capability to spread diseases, and a list of subzones which were more vulnerable in terms of being a receiver. In summary, the identified subzones were mainly the core area of residential and transportation hubs. This places had high population activity, such as transportation hubs or community hubs. Therefore, these places should be more important for the resource allocation and disease monitoring for the prevention and intervention purpose. For example, the public health agencies could consider these locations while planning to setup the body temperature sensing equipment, to provide personal hygiene toolkit, or while setting up the advertisements for reminders or related information. On the other hand, since these locations were more vulnerable and more influential, they should get more attentions while setting up the policies such as the temporary closure of some businesses or restrictions of large scale activities.

## Conclusion

This study provided a network analysis framework to consider the local flow intensity and also the neighborhood diversity of zones and coreness to measure the ability to spread or receive some entities. From the theoretical perspective, the framework in this study considered weighted and directed interactions between nodes (places) to identify super-spreaders and super-receivers. From the practical perspective, this study demonstrated an analysis of the identification of the influential and vulnerable locations based on the public transport flow network. This study used Singapore public transport flow network as a case study to uncover the important locations of disease diffusion.

There are several limitation in this study. First, this study covered only the public transportation commuters, specifically, only bus and train riderships were included. Other ways of transportation, including the private or hired automobiles (cars, motorcycles, shuttle buses or vans), and active transportation (by walking, bicycles, skateboards, scooters, etc.) were not included. Some of the subzones currently did not

have bus stops or train stations. As aforementioned the public transportation were the major travelling modes, i.e. covered more than 60% of daily commuting, we believed that the results from this study were adequate to capture the big picture of the human movement patterns. Second, Singapore is an island country with its North national border connected to Malaysia through two land checkpoints. But the cross-border flows were not included in this study. Many workers and students commute between Singapore and Malaysia in daily basis. There are some bus services directly connecting stations in Johor Bahru, Malaysia and various places across Singapore, including Woodlands at the North Region, Jurong East at the West Region, and Bugis at the Central Region, etc. Since these data were ignored, the in/out flows of these places in Singapore would be underestimated.

Third, inter-mode trip transfer and bus transfer were not considered in the datasets. The trip transfer between MRT lines were captured from the tap-in and tap-out records, i.e. passengers exchange lines within train stations. But the OD data for buses only records the direct flow between bus stops, i.e. the records present only the tap-in and tap-out a bus information, the records of the exchange of bus services were not shown in the data. On the other hand, the data about changing from bus to train and vice versa were also unavailable. Therefore, we can only capture the direct bus service and this would limits the movement of travelers to the existence of direct bus services or train services. Forth, the dynamic through a day was ignored. This study aggregated all hours data together to reveal the daily basis commuting flow network pattern. The within day dynamic could be different and the super-spreader and super-receiver could be different. The dynamic of spreader and receiver indexes could be done in a future study.

In this study, we had developed a framework to identify spatial super-spreaders and super-receivers. We believed that there were two directions of studies could further extend the concept and usage of this study in future researches. The first direction were to test and compare the super-spreaders and super-receivers using some disease diffusion simulation models for spatial networks. The dynamic patterns in the disease diffusion could be observed from the simulation models, and thus the effect of the super-spreaders and receivers could be quantified through these models. Second, the geography, demography, and social-economic of the spatial super-spreaders and super-receivers could be analysed using some statistical models, to identify the potential social and physical environmental factors that made these locations a super-spreaders or receivers.

## Acknowledgments

Cras egestas velit mauris, eu mollis turpis pellentesque sit amet. Interdum et malesuada fames ac ante ipsum primis in faucibus. Nam id pretium nisi. Sed ac quam id nisi malesuada congue. Sed interdum aliquet augue, at pellentesque quam rhoncus vitae.

## References

1. WHO. Coronavirus Disease 2019 (COVID-19) Situation Report 40. Switzerland: WHO; 2020. 40.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet*. 2020;395(10224):565–574.  
doi:10.1016/S0140-6736(20)30251-8.

3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497–506. doi:10.1016/S0140-6736(20)30183-5.
4. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, et al. Epidemiological and Clinical Features of the 2019 Novel Coronavirus Outbreak in China. *Epidemiology*; 2020.
5. Riou J, Althaus CL. Pattern of Early Human-to-Human Transmission of Wuhan 2019 Novel Coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*. 2020;25(4). doi:10.2807/1560-7917.ES.2020.25.4.2000058.
6. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. 2020; p. NEJMoa2001316. doi:10.1056/NEJMoa2001316.
7. WHO. Novel Coronavirus(2019-nCoV) Situation Report 12. Switzerland: WHO; 2020. 12.
8. WHO. Novel Coronavirus(2019-nCoV) Situation Report 9. Switzerland: WHO; 2020. 9.
9. WHO. Coronavirus Disease 2019 (COVID-19) Situation Report 59. Switzerland: WHO; 2020. 59.
10. WHO. Coronavirus Disease 2019 (COVID-19) Situation Report 47. Switzerland: WHO; 2020. 47.
11. Pastor-Satorras R, Vespignani A. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*. 2001;86(14):3200–3203. doi:10.1103/PhysRevLett.86.3200.
12. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of Influential Spreaders in Complex Networks. *Nature Physics*. 2010;6(11):888–893. doi:10.1038/nphys1746.
13. Fu YH, Huang CY, Sun CT. Identifying Super-Spreader Nodes in Complex Networks. *Mathematical Problems in Engineering*. 2015;2015:1–8. doi:10.1155/2015/675713.
14. Liu HL, Ma C, Xiang BB, Tang M, Zhang HF. Identifying Multiple Influential Spreaders Based on Generalized Closeness Centrality. *Physica A: Statistical Mechanics and its Applications*. 2018;492:2237–2248. doi:10.1016/j.physa.2017.11.138.
15. Stein RA. Super-Spreaders in Infectious Diseases. *International Journal of Infectious Diseases*. 2011;15(8):e510–e513. doi:10.1016/j.ijid.2010.06.020.
16. Edholm CJ, Emerenini BO, Murillo AL, Saucedo O, Shakiba N, Wang X, et al. Searching for Superspreaders: Identifying Epidemic Patterns Associated with Superspreading Events in Stochastic Models. In: Radunskaya A, Segal R, Shtylla B, editors. *Understanding Complex Biological Systems with Mathematics*. vol. 14. Cham: Springer International Publishing; 2018. p. 1–29.
17. Liu JG, Ren ZM, Guo Q. Ranking the Spreading Influence in Complex Networks. *Physica A: Statistical Mechanics and its Applications*. 2013;392(18):4154–4159. doi:10.1016/j.physa.2013.04.037.

18. Zeng A, Zhang CJ. Ranking Spreaders by Decomposing Complex Networks. *Physics Letters A*. 2013;377(14):1031–1035. doi:10.1016/j.physleta.2013.02.039.
19. He JL, Fu Y, Chen DB. A Novel Top-k Strategy for Influence Maximization in Complex Networks with Community Structure. *PLOS ONE*. 2015;10(12):e0145283. doi:10.1371/journal.pone.0145283.
20. Wang X, Zhang X, Zhao C, Yi D. Maximizing the Spread of Influence via Generalized Degree Discount. *PLOS ONE*. 2016;11(10):e0164393. doi:10.1371/journal.pone.0164393.
21. Chen D, Lü L, Shang MS, Zhang YC, Zhou T. Identifying Influential Nodes in Complex Networks. *Physica A: Statistical Mechanics and its Applications*. 2012;391(4):1777–1787. doi:10.1016/j.physa.2011.09.017.
22. Gao S, Ma J, Chen Z, Wang G, Xing C. Ranking the Spreading Ability of Nodes in Complex Networks Based on Local Structure. *Physica A: Statistical Mechanics and its Applications*. 2014;403:130–147. doi:10.1016/j.physa.2014.02.032.
23. Li C, Wang L, Sun S, Xia C. Identification of Influential Spreaders Based on Classified Neighbors in Real-World Complex Networks. *Applied Mathematics and Computation*. 2018;320:512–523. doi:10.1016/j.amc.2017.10.001.
24. Zhang X, Zhu J, Wang Q, Zhao H. Identifying Influential Nodes in Complex Networks with Community Structure. *Knowledge-Based Systems*. 2013;42:74–84. doi:10.1016/j.knosys.2013.01.017.
25. Barthélémy M. Spatial Networks. *Physics Reports*. 2011;499(1-3):1–101. doi:10.1016/j.physrep.2010.11.002.
26. Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, Kong J, et al. Understanding the Spatial Clustering of Severe Acute Respiratory Syndrome (SARS) in Hong Kong. *Environmental Health Perspectives*. 2004;112(15):1550–1556. doi:10.1289/ehp.7117.
27. Colizza V, Vespignani A. Epidemic Modeling in Metapopulation Systems with Heterogeneous Coupling Pattern: Theory and Simulations. *Journal of Theoretical Biology*. 2008;251(3):450–467. doi:10.1016/j.jtbi.2007.11.028.
28. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the Spatial Spread of Infectious Diseases: The Global Epidemic and Mobility Computational Model. *Journal of Computational Science*. 2010;1(3):132–145. doi:10.1016/j.jocs.2010.07.002.
29. Chin WCB, Wen TH, Sabel CE, Wang IH. A Geo-Computational Algorithm for Exploring the Structure of Diffusion Progression in Time and Space. *Scientific Reports*. 2017;7(1):12565. doi:10.1038/s41598-017-12852-z.
30. Hsieh YH, van den Driessche P, Wang L. Impact of Travel Between Patches for Spatial Spread of Disease. *Bulletin of Mathematical Biology*. 2007;69(4):1355–1375. doi:10.1007/s11538-006-9169-6.
31. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al. The Role of Human Movement in the Transmission of Vector-Borne Pathogens. *PLoS Neglected Tropical Diseases*. 2009;3(7):e481. doi:10.1371/journal.pntd.0000481.

32. Nicolaides C, Cueto-Felgueroso L, González MC, Juanes R. A Metric of Influential Spreading during Contagion Dynamics through the Air Transportation Network. *PLoS ONE*. 2012;7(7):e40961. doi:10.1371/journal.pone.0040961.
33. Jiang B. Ranking Spaces for Predicting Human Movement in an Urban Environment. *International Journal of Geographical Information Science*. 2009;23(7):823–837. doi:10.1080/13658810802022822.
34. Zhong C, Arisona SM, Huang X, Batty M, Schmitt G. Detecting the Dynamics of Urban Structure through Spatial Network Analysis. *International Journal of Geographical Information Science*. 2014;28(11):2178–2199. doi:10.1080/13658816.2014.914521.
35. Chin WCB, Wen TH. Geographically Modified PageRank Algorithms: Identifying the Spatial Concentration of Human Movement in a Geospatial Network. *PLOS ONE*. 2015;10(10):e0139509. doi:10.1371/journal.pone.0139509.
36. Meloni S, Perra N, Arenas A, Gómez S, Moreno Y, Vespignani A. Modeling Human Mobility Responses to the Large-Scale Spreading of Infectious Diseases. *Scientific Reports*. 2011;1(1):62. doi:10.1038/srep00062.
37. Aral S, Walker D. Identifying Influential and Susceptible Members of Social Networks. *Science*. 2012;337(6092):337–341. doi:10.1126/science.1215842.
38. Moore C, Cumming GS, Slingsby J, Grewar J. Tracking Socioeconomic Vulnerability Using Network Analysis: Insights from an Avian Influenza Outbreak in an Ostrich Production Network. *PLoS ONE*. 2014;9(1):e86973. doi:10.1371/journal.pone.0086973.
39. Porphyre T, Correia-Gomes C, Chase-Topping ME, Gamado K, Auty HK, Hutchinson I, et al. Vulnerability of the British Swine Industry to Classical Swine Fever. *Scientific Reports*. 2017;7(1):42992. doi:10.1038/srep42992.
40. Dhewantara PW, Mamun AA, Zhang WY, Yin WW, Ding F, Guo D, et al. Geographical and Temporal Distribution of the Residual Clusters of Human Leptospirosis in China, 2005–2016. *Scientific Reports*. 2018;8(1):16650. doi:10.1038/s41598-018-35074-3.
41. Kleinberg JM. Hubs, Authorities, and Communities. *ACM Computing Surveys (CSUR)*. 1999;31(4es):5. doi:10.1145/345966.345982.
42. Rosvall M, Axelsson D, Bergstrom CT. The Map Equation. *The European Physical Journal Special Topics*. 2009;178(1):13–23. doi:10.1140/epjst/e2010-01179-1.
43. Garas A, Schweitzer F, Havlin S. A k -Shell Decomposition Method for Weighted Networks. *New Journal of Physics*. 2012;14(8):083030. doi:10.1088/1367-2630/14/8/083030.
44. Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E. A Model of Internet Topology Using K-Shell Decomposition. *Proceedings of the National Academy of Sciences*. 2007;104(27):11150–11154. doi:10.1073/pnas.0701175104.
45. Pei S, Muchnik L, Andrade JS, Zheng Z, Makse HA. Searching for Superspreaders of Information in Real-World Social Media. *Scientific Reports*. 2015;4(1). doi:10.1038/srep05547.

46. Garske T, Rhodes CJ. The Effect of Superspreading on Epidemic Outbreak Size Distributions. *Journal of Theoretical Biology*. 2008;253(2):228–237.  
doi:10.1016/j.jtbi.2008.02.038.