

Identifying Critical Components of a Public Transit System for Outbreak Control

András Bóta¹ · Lauren M. Gardner¹ ⁽¹⁾ · Alireza Khani²

Published online: 8 August 2017 © Springer Science+Business Media, LLC 2017

Abstract Modern public transport networks provide an efficient medium for the spread of infectious diseases within a region. The ability to identify components of the public transit system most likely to be carrying infected individuals during an outbreak is critical for public health authorities to be able to plan for outbreaks, and control their spread. In this study we propose a novel network structure, denoted as the vehicle trip network, to capture the dynamic public transit ridership patterns in a compact form, and illustrate how it can be used for efficient detection of the high risk network components. We evaluate a range of network-based statistics for the vehicle trip network, and validate their ability to identify the routes and individual vehicles most likely to spread infection using simulated epidemic scenarios. A variety of outbreak scenarios are simulated, which vary by their set of initially infected individuals and disease parameters. Results from a case study using the public transit network from Twin Cities, MN are presented. The results indicate that the set of transit vehicle trips at highest risk of infection can be efficiently identified, and are relatively robust to the initial conditions of the outbreak. Furthermore, the methods are illustrated to

Lauren M. Gardner l.gardner@unsw.edu.hu

> András Bóta a.bota@unsw.edu.au

Alireza Khani akhani@umn.edu

- Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, University of New South Wales (UNSW) Australia, Sydney NSW 2052, Australia
- ² Department of Civil, Environmental and Geo-Engineering, University of Minnesota Twin Cities, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA

be robust to two types of data uncertainty, those being passenger infection levels and travel patterns of the passengers.

Keywords Network modeling \cdot Public transportation \cdot Infection models \cdot Outbreak control \cdot Public health

1 Introduction

In many parts of the world a large proportion of the population lives and commutes in increasingly dense conditions, ideal for rapid disease transmission. Of particular interest in this study is the role of the public transit system during an outbreak, which can act as a potential catalyst in the disease transmission process within a metropolitan region. As an example, empirical studies based on the SARS outbreak in Beijing illustrated public transit usage corresponded to an increased risk of infection (Wu et al. 2004; Troko et al. 2011). However, due to the stochastic nature of the disease spreading process (i.e., contact between an infectious and susceptible person may or may not result in a new infection) and the lack of information traditionally available on individuals' daily contact patterns, it is extremely difficult to predict the impact a new disease might have on a region, and subsequently plan for and control infectious disease outbreaks.

One approach to improving outbreak management is through identification of "hot spot locations" which should be closely monitored and potentially targeted for outbreak control in real time. Thus, the goal of this study is to identify such potential vulnerabilities in the regional public transit system in the event of an outbreak. To do so, we proposed a novel network structure to represent the contacts and movement of individuals using the public transit system, which can be used to efficiently identify critical components of the system (e.g., super spreading vehicle-trips) (Galvani and May 2005; Kitsak et al. 2010). Outputs from an activity-based travel demand model for Twin Cities, MN is used to generate the network structure, and results for a case study are presented.

2 Background

The spread of contact-based infectious diseases can be modeled as a dynamic process atop social contact networks (Murray 2002; Anderson and May 1991). Using this approach, various agent based simulation models have been developed and implemented to replicate possible spreading scenarios, predict average spreading behavior, and analyze various intervention strategies for a given network and disease (Fajardo and Gardner 2013; Gardner et al. 2014; Rey et al. 2016; Rvachev and Longini 1985; Epstein et al. 2002; Eubank 2004; Hufnagel et al. 2004; Dibble and Feldman 2004; Cahill et al. 2005; Dunham 2005; Meyers et al. 2005; Small and Tse 2005; Carley et al. 2006; Ferguson 2005; Germann et al. 2006). Similar models have also been used to model the propagation of information, ideas and opinions between individuals (Ekici et al. 2008; Roche et al. 2011).

To accurately map disease spread, models must exploit characteristics of the population and the disease itself; and thus require large amounts of data, and significant computational resources. However, computational capabilities are exponentially increasing, and developments in data collection are increasingly providing a means for accurate mappings between known individuals. In particular, spatial analysis of transport and communication networks which exploits these data sources are a growing area of research (Haydon et al. 2003; Coleman et al. 1996; Hasan and Ukkusuri 2011; Candia et al. 2008; Gastner and Newman 2006; Schintler et al. 2007; Erath et al. 2009; Wang et al. 2009; González et al. 2006, 2008). The ongoing development of activity-based travel models, which examine why, where and when various activities are engaged in by individuals (Brockmann et al. 2006; Song et al. 2010; De Montjoye et al. 2013; Chen et al. 2016; Gardner et al. 2012), as well as innovations in pedestrian modeling (Gardner and Sarkar 2013) also present additional promising alternatives to generate social-contact networks in the future. These efforts have recently been expanded to include social network modeling, specifically the ability to reproduce spatial structure and interaction between individuals for large-scale social networks (Chen et al. 2016; Gardner et al. 2012; Lam and Huang 2003; Roorda et al. 2009).

Of particular applicability to this study are the recent advances in public transit modeling, which can now provide detailed contact patterns including temporal patterns (e.g. bus travel time), and spatial patterns (a function of the vehicle size and passenger volume) (Ramadurai and Ukkusuri 2010; Illenberger et al. 2012). While these methods can potentially allow accurate mappings between known individuals, the data collection and processing required to recreate a real-world contact network is expensive, computationally costly, and time intensive, among other challenges faced such as privacy issues (Huerta and Tsimring 2002; Hoogendoorn and Bovy 2005; Balcan 2009; Salathé 2010; Funk et al. 2010; Nassir et al. 2012). As such, it is critical to develop methods which can exploit this data as it becomes available.

Such examples have been recently addressed in the literature. Pendyala et al. (2012) highlights the value of mobile network data for modeling human mobility patterns in real time, the results of which can be used in designing surveillance and containment strategies during an outbreak. The authors constructed intra and international mobility patterns in the set of African countries affected by and surrounding the Ebola outbreak. Their work exemplifies the role of increasingly available real-time connectivity data in public health response. Similarly, Sun et al. (2013) generated a physical encounter network using daily transit data from Singapore smart cards, and then used the contact network generated to compare different individual-based sensor schemes for early detection of simulated outbreaks (Christakis and Fowler 2010). Their work utilized a "friends as sensors" approach, similar to that in Cattuto (2010), Stehlé (2011), Kuiken et al. (2000), Gilbert (2007), Wesolowski et al. (2014), applied atop a more complete physical contact network. As an alternative to focusing on the most likely infected individuals, Bóta et al. (2017) proposed a method to identify the components of the public transit infrastructure at highest risk of furthering the spread of infection. This was accomplished by first simulating various outbreak scenarios atop a generated passenger contact network, and second, translating the simulation results to identify the bus trips most likely to carry infected passengers.

Similarly to Bóta et al. (2017), this study seeks to identify the highest risk vehicles in a public transit system in the event of an outbreak, but introduces an entirely different method to do so. The contribution in this work is the development of a novel network structure that is capable of capturing the dynamics of passenger travel in a compact form, which can be used to efficiently identify the highest risk components of a public transit system, i.e., the trips most likely to further the spread of disease. Thus, the proposed method does not rely on running an infection simulation model, or on a complete contact graph. We denote the new structure as the vehicle trip network, where a vehicle trip refers to a bus or train route with a specific departure time.

The analysis is conducted in four stages. In the first stage we generate a passenger contact network using the output from an activity-based travel demand model. Some general properties of the contact network are presented and discussed. The second stage of analysis focuses on modeling epidemic outbreaks on the contact network generated. To do this an agent based stochastic simulation model is used to model the spread of infection between individuals based on contact patterns. A range of outbreak scenarios, which vary based on the number and location of initially infected individuals as well as disease characteristics, e.g., infectiousness, are simulated atop the contact network, and the resulting infection patterns (i.e., set of passengers most likely to be infected) are evaluated. The set of initially infected nodes are selected using both random and targeted selection strategies, representing both naturally occurring and malicious introductions of infection into the system, respectively. The passenger-level outbreak results are also interpreted to estimate the risk posed at the vehicle level, i.e., the set of vehicle trips most likely to carry infected passengers. In the third stage of the analysis the vehicle trip network is created using two different approaches. In the fourth stage we seek to best replicate the outbreak behavior revealed in stage two, i.e., the set of highest risk vehicle trips, but now using only the vehicle trip network. This is accomplished by computing several network-based statistics on the vehicle trip network directly, and is shown to be capable of accurately identifying the majority of high risk trips. In order to model the uncertainty, that is usually encountered when collecting real-world data, we introduce noise to both second and fourth stages of the analysis, and show that our method is robust to such errors. Additionally, the size of the vehicle trip network is significantly smaller than the contact network, and the metrics are significantly easier to compute than running the stochastic simulation model multiple times, therefore the new network structure provides a substantial computational advantage.

The remainder of this manuscript includes a description of the data used in the study, the methodology implemented, followed by numerical results and conclusions. Results from a case study using the public transit network from Twin Cities, MN are presented.

3 Data

Public transportation data in this study were obtained from the transit system in Twin Cities region in Minnesota, where 187 routes serve 13,700 stops in the region. Transit network and schedule data were created from General Transit Feed Specification

(GTFS), including near 0.5 Million stop-times on a weekday in 2015. Transit passenger trips were obtained from Metropolitan Councils's activity-based demand model, and contained more than 293,000 linked trips (i.e. a passenger trip from an origin to a destination that may include zero or more transfers). A transit demand table was then assigned to the transit network based on a logit route choice model, and passengers were simulated in the transit network using a mesoscopic transit passenger simulation model. These steps were done using the FAST-TrIPs model (Khani et al. 2015). The output of the transit assignment model used in this paper is the travel path of each individual passenger, that is for each passenger p, we know i) identifiers of the vehicles p traveled on, ii) the time of boarding for all involved vehicles and iii) the time of alighting for all involved vehicles. Note, that the information collected this way can be obtained from smart card systems, offering good adaptability for the proposed method.

4 Network Structures

The collected transit data naturally defines a network structure. We denote this network as the *contact network*. The nodes of this network are passengers and a link exists between them if they were traveling on the same vehicle. Since this relationship is mutual, the network is undirected. The time dimension of the network is defined on the links of the network by two values: a *time stamp* on a link indicates the start of the contact between the two incident passengers and a *contact duration* value indicates the length of the contact.

An additional attribute is available on the edges of the contact network, the id of the vehicle trip where the contact takes place. We define the *vehicle trip* as the trip a single vehicle takes on a specific route with a specific departure time. This also tells us which vehicle trips the passengers travelled on; that is we have a set of vehicle trips assigned to each passenger, and we also have the set of passengers for each vehicle trip. We denote the passengers travelling on multiple trips as *transfer passengers*, while the passengers with only a single vehicle trip will be called *bulk passengers*.

4.1 Contact Network

The contact network corresponding to the Twin Cities, MN dataset has 94475 passengers and 6287847 contacts between them. The density plot of the contact start times is illustrated in Fig. 1a. The start time distribution has two peaks, one around 7 AM and another around 5 PM, which is consistent with a morning and evening weekday commute. The degree distribution of the network is shown in Fig. 1b. It follows a skewed power law, where the average number of contact per person is 136 while the maximum is 827. A large majority (97%) of the passengers ride more than one vehicle on a single workday; 38% of the passengers ride exactly two vehicles, while 16% ride more than four.

Since the network is too large to be appropriately visualized, we show two typical subgraphs in Fig. 2. In both figures the nodes represent individual passengers and the links connect passengers that ride on the same vehicle trip. Figure 2a illustrates a



Fig. 1 The distribution of a contact start times and b the degree distribution of the contact network

passenger with a high number of contacts, represented by black node in the middle, and links to all other contacts (1-neighbors) in its direct neighborhood. The nodes are colored according to vehicle trips, and darker edges indicate longer contact durations. Nodes are grouped according to vehicle trips. In reality, many of these nodes were present on multiple vehicle trips, and the color only indicates the first vehicle trip with the central passenger.

Figure 2b illustrates the contact network of the vehicle trip with the highest ridership. Here, the nodes are colored according to the contact start time, with lighter



Fig. 2 Two subgraphs of the contact network. **a** The neighborhood of a single passenger with a high number of contacts. The passenger is represented by the *black dot* in the *middle*. The *colors* are indicative of the vehicle trips the passengers first met on, with darker edges representing longer contact durations. **b** illustrates the contact network of the most frequented vehicle trip where colors indicate contact start times

colors corresponding to earlier connections. The links are shaded according to contact duration. This figure depicts how contacts on a single trip change in time. It reveals that the passengers that meet earlier in the morning tend to travel together for longer durations. It should be noted that not all of the passengers are on the vehicle at the same time.

4.2 Vehicle Trip Network

The novelty of this work lies in the development and analysis of the proposed vehicle trip network. The task of the vehicle trip network is to characterize the movement of passengers using a public transportation system in a compact way. Below, we propose two methods to construct the vehicle trip network, each of them represents a different kind of interaction mechanics among vehicle trips.

We define a *vehicle trip* as the trip a single vehicle takes on a specific route with a specific departure time. Given the information from the contact network we know which vehicle trips the passengers traveled on; that is, we have a set of vehicle trips assigned to each passenger, and we also have the set of passengers for each vehicle trip. We denote the passengers traveling on multiple trips as *transfer passengers*, while the passengers with only a single vehicle trip will be called *bulk passengers*.

In both vehicle trip networks the nodes represent particular vehicle trips, and links connect vehicle trips if there is a transfer of passengers between them. The transfer of passengers has a natural direction, always from a vehicle trip to another, making the vehicle trip network directed. The links are weighted based on the number of passengers transferring from one vehicle trip to the other. The nodes and the links of the vehicle trip network therefore correspond to node sets of the contact network. Both approaches create a structure that is very similar to the intersection graph formed by the passenger groups corresponding to each individual vehicle trip. In the analysis section we evaluate both methods, and discuss the benefits of each.

Approach one

The contact start times on the links of the contact network can be used to reconstruct the travel route of each passenger. We define this route as the *passenger path*, which is an ordered list of vehicle trips the passenger was traveling on. Based on the ordering of this list an index number is assigned to each vehicle trip in the list that is specific to both the vehicle trip and the list, i.e. trips appearing in multiple passenger paths have an index in each path they are present in.

In the first approach to construct the vehicle trip network we use all passenger paths in the following way: the nodes of the vehicle trip network are the vehicle trips, and a directed link exists between two of them if they appear in subsequent positions in any passenger path. That is, if u and w are vehicle trips, there is a link between them if there is a passenger path P such that $u, w \in P$, and $i_P(u) = i_P(w) + 1$, or $i_P(w) = i_P(u) + 1$, where $i_P(v)$ denotes the index of vehicle trip v in passenger path P. For example considering the passenger path $p_p = \{v_1, v_2, v_3, v_4\}$, three links are created: $e_1(v_1, v_2), e_2(v_2, v_3), e_3(v_3, v_4)$. We can assign a weight value to the links depending on how many times the two vehicles appeared in succeeding positions across all passengers.

Approach two

In the second approach we again construct the vehicle trip network using all passenger paths, but extend the definition of a link. In approach two, the nodes of the vehicle trip network are again the vehicle trips, but we add a directed link connecting vehicle trips u and w if:

- 1. they appear in subsequent positions in any passenger path (as in approach 1) or
- 2. $u, w \in P$ and $i_P(u) = 1, i_P(w) \ge 3$ or $i_P(w) = 1, i_P(u) \ge 3$.

In the same example passenger path $p = v_1, v_2, v_3, v_4$, using approach two, five links would be created, whereas in approach one only three links are created, $e_1(v_1, v_2), e_2(v_2, v_3), e_3(v_3, v_4)$. The additional two links, $e_4(v_1, v_3)$ and $e_5(v_1, v_4)$ would be generated according to the second criterion above. This criterion allows the connection of all pairs of vehicle trips which appear in a passengers path, irrespective of the order. Since infection always spreads from v_1 to other vertices, there is no link between v_2 and v_4 . In approach two, weights are assigned to a link connecting two vehicle trips based on the total number of passengers that ride on both vehicle trips (at any point in their path).

Examples of both approaches can be seen on Fig. 3. There are a total of 12 passengers traveling on six vehicle trips. Figure 3a shows the contact network, where the nodes are grouped according to the vehicle trips. Figure 3b shows the vehicle trip network created with approach one, while 3/c shown the network created with approach two. The passengers traveling on each trip and the passengers transferring from one trip to another are visible. The order in which passengers transfer from trip to trip may change the structure of the network. In this example, passenger G traveled on trips 2, 3 and 5 in this order, hence the direction of the links on Fig. 3b and c.

The differences between the two approaches are subtle but important for finding a relation between the behavior of contact and vehicle trip networks. Approach one is a subgraph of approach two, while approach two contains additional links not included in approach one. To see why the additional links are desirable, consider passenger G in the example on Fig. 3. Passenger G travels on three vehicle trips: 2, 3 and 5. Under approach one, there is no direct interaction between trips 2 and 5 even though G was present on both of them. The vehicle trip network created with approach two has a link between 2 and 5, and thus allows the interaction. Now, consider a hypothetical infection scenario, where passenger K is infected with an infectious disease. In this instance the infection could spread from trip 2 to both trips 3 and 5 in a single day. In approach one the link between trip 2 and 5 is missing, disallowing a potential infection event, while in approach two it is possible to capture on the resultant network. Because our goal in this paper is to accurately and efficiently model disease spreading risk within the transit ridership network, representing all possible paths of infection is critical. We will show in Section 7, that the additional links provided by approach two noticeably improve the accuracy of the model.

Compared with the passenger-level contact network, the proposed vehicle trip network represents a much more efficient network structure. For the Twin Cities, MN dataset, the contact network has 94475 nodes (i.e., passengers) and 6287847 edges (i.e., contacts), where the vehicle trip networks have 8002 nodes (i.e., vehicle trips)



Fig. 3 Relationship between contact and vehicle trip networks. \mathbf{a} is an example of a contact network, the ellipses indicate passengers traveling on the same vehicle trip. \mathbf{b} is the corresponding vehicle trip network created with approach one, \mathbf{c} with approach two. The letters inside the nodes denote the passengers traveling on the trip, while the letters on the links denote the passengers transferring from one trip to another

and 164677 and 263792 edges for approaches one and two, respectively. This significant reduction in network size offers a potentially much more efficient means to evaluate the network-level risk posed by passengers. The following analysis reveals that the vehicle trip network is able to capture the dynamics of the public transit ridership patterns and vehicle trip interactions, and can be used to accurately identify the critical components of the system in the event of an outbreak.

5 Modeling the Spread of Infection

In the first part of our analysis we model a variety of infection scenarios on the contact network. The network-based infection model described in the following section is used to model the spread of disease between the individuals. We consider multiple infection scenarios differing in both the size and spatial distribution of the initially infected individuals, and the level of infectiousness of the disease. We model the spread of an infectious disease over a five day period, where the contact network remains constant day-to-day.

5.1 Infection Simulation Model

A discrete compartmental susceptible-infected-removed (SIR) model is used to model an epidemic outbreak on the network, and is defined as follows. The SIR infection process takes place on a network defined as G(V, E), where V(G) denotes the node set of G, and E(G) denotes the edge set of G. The process can be defined on directed and undirected networks, but in this study we only consider the latter case. Real values denoted as edge infection probabilities are required on the edges of the network, these are defined as $w_e \in [0, 1]$ for all $e \in E(G)$. An edge infection probability w_e on an edge e(u, v) denotes the probability that if node u is infected, the infection spreads to node v, or if v is infected, the infection spreads to u.

During the SIR infection process all nodes are required to adopt one of the three available states. A node can be susceptible (S), meaning it is not infected and can become infected, infected (I), meaning it is infected and may spread the disease to neighboring nodes, or removed (R) which means it was previously infected and has since recovered and is no longer able to spread infection. A discrete time period is attached to the infected state denoted as τ_i indicating the amount of time a node spends in this state. The non-empty set of initially infected nodes will be denoted as $A_0 \subset V(G)$, the nodes are the sources of the infection. We define an infection scenario as the triple $S(G, W, A_0)$, where G is a graph, $W: E(G) \mapsto [0, 1]$ is a surjective assignment of edge infection probabilities to the edges of the graph, and A_0 denotes the set of initially infected nodes. The infection scenario defines the input of an infection process.

The infection process itself is iterative and takes place in discrete time steps, according to the following logic. Let $A_i \subseteq V(G)$ denote the set of infected nodes in iteration *i*. Each infected node $u \in A_i$ tries to infect its susceptible neighbors $v \in V \setminus \bigcup_{0 \le j \le i} A_j$ according to w_e . If the attempt is successful *v* will be in an infected state starting from iteration i + 1. If the attempt is unsuccessful, node *u* may make additional attempts to infect *v* depending on *i* and τ_i . If more than one node is trying to infect *v* in the same iteration, the attempts are made in an arbitrary order independent of each other. Finally all nodes in an infected state change their state to removed if the time they became infected t_u is $t_u = i - \tau_i$.

The application of the SIR infection model to the contact network in this study can be interpreted as follows: Each iteration corresponds to a single day. If a person gets infected in iteration *i* then he is infectious for τ_i days, starting from iteration i + 1, indicating a latency period of a single day. In this study we assume an infectious period (τ_i) of five days. Results from the process provide the state of each individual (S, I or R) and each iteration. Due to the characteristics of the dataset we consider a five day observation period; therefore, any infected individual remains infectious throughout the entire simulation (i.e., there will be no recovered individuals within the time-frame considered).

Due to the stochastic nature of the model, we seek to estimate the likelihood of infection for each node at each iteration. In this study for each outbreak scenario considered we run the infection process k times and count the frequency of infection for each node at any given iteration to estimate the probability that a node is in an infectious state at any iteration. For all results in this paper k = 10,000 was used.

This approach is similar to that used in Bóta et al. (2013), Kempe et al. (2003), Bóta et al. (2017).

5.2 Experiment Design

Given the stochastic nature of infection spread, we consider a diverse set of infection scenarios on the contact network. The scenarios represent a high variety of initial outbreak conditions, which can be divided into two strategies according to the definition of the infection scenario in the previous section. The edge infection probability assignment represents the level of infectiousness of a particular disease, while the set of initially infected nodes represents the source(s) of the outbreak. The set of initially infected nodes can be further divided into two factors: the size of the initial set and the spatial distribution of it.

Following the assumption of Sun et al. (2013) we define the transmission probabilities as a linear function of the contact duration between individuals: $p(u, v) = d(u, v) * \beta$, where p(u, v) is the transmission probability between nodes u and v, d(u, v) is the contact duration between u and v in minutes, and β is a constant. We consider four different values for β which represent a wide range of outbreaks; $\beta = 0.0005$ puts the edge infection values between 0 and 0.05, $\beta = 0.001$ puts the transmission probabilities between 0 and 0.1, $\beta = 0.0015$ with the bounds 0 and 0.15 and $\beta = 0.002$ with 0 and 0.2.

The size of the initially infected node set A_0 will be 10, 50 and 100 representing a small, medium and large number of initially infected individuals. These individuals are selected according to two different strategies, similar to the works of Albert and Barabási (2002). The first strategy considers the top 10, 50 and 100 most central individuals according to their degree centrality corresponding to a "targeted" strategy, while the second strategy distributes the same amount of nodes randomly. In the random selection strategy ten samples were drawn independently from V(G) for each scenario, and the test results were averaged.

5.3 Limitations of the Study

Before presenting the results certain limitations of this study should be noted. The first set of limitations is due to the available data. The contact network is generated using a single day of data, but used to model the spread of an outbreak over a 5-day period. This implies the assumption that commute patterns remain constant day-to-day over a 5-day work week. While individual's travel patterns can vary day-to-day, a recent study (Sun et al. 2013) which used travel smart card data to generate an in-vehicle social encounter network on public buses using a full week of travel data found that physical encounters display reproducible temporal patterns. The finding that repeated encounters are regular and identical, and rooted in daily behavioral regularity supports the assumption used in this work.

Implicit assumptions also result from the use of the SIR model, which restrict this study to the family of infectious diseases that are transmitted from an infected to susceptible individual via direct or close contact. This category includes various strands of the flu, SARS and the common cold, among others. These assumptions could

easily be relaxed without altering the proposed methodology, and a more complex compartmental model, e.g., SEIR, could be substituted.

Finally, this study ignores all contacts made outside of transit movements, which in reality would significantly increase the size of the outbreak within a region (while also resulting in more infected transit riders). However, predicting the size of the outbreak is not the intended goal of this study. Instead, the contribution is the proposed methodology to identify the components of the public transit system that play a critical role during the early stages of an outbreak. The identified transit-trips can be prioritized for surveillance monitoring and possible mitigation and control efforts by transit and public health authorities at the initial stages of an outbreak.

6 Contact Network Analysis

The following analysis reveals the outbreak spreading behavior on the contact network. Our key point of interest is the behavior as a function of the initial conditions, e.g., number and set of initially infected individuals. Apart from the fraction of infected individuals we are most interested in the components of the transit system that are at highest risk of infection. We present all results as of day 5; however analogous results for any stage of the outbreak are available.

6.1 Outbreak Scenario Evaluation

The Twin Cities, MN dataset, which has 94475 passengers and 6287847 contacts between them, is a large network. As mentioned in the previous section, an SIR simulation was implemented and run on this network with a frequency parameter k = 10,000.

Table 1 reveals the fraction of infected individuals in the contact network (on day 5) for each infection scenario considered. The rows correspond to the transmission probabilities, while the columns indicate the size of the initially infected set and the selection strategy. As expected, the prevalence of the outbreak increases with number of initially infected individuals, i.e., size of A_0 , and the level of infectiousness of the disease β . Results are also provided for the targeted and random strategies, which only vary based on the spatial distribution of the initial set of infected individuals. Computational time varied greatly depending on the infection scenario ranging from 26 seconds to 72 minutes. Scenarios with a high observed number of infected passengers took more time to compute.

The results in Table 1 generally confirm previous observations (Albert and Barabási 2002) that a targeted strategy has a much greater impact on the network, than a random one. This is true for all infection scenarios presented here, although the difference is typically less when the network is highly infected. The standard deviation values for the random selection strategy (indicated by the values in parenthesis in Table 1) also vary significantly across scenarios. These represent how the expected size of the infection values of the less infectious scenarios (small β and $|A_0|$) values reveal that the selection of the initially infected node set is critical if the

	Targeted scenarios			Random scenarios		
	$ A_0 = 10$	$ A_0 = 50$	$ A_0 = 100$	$ A_0 = 10$	$ A_0 = 50$	$ A_0 = 100$
$\beta = 0.0005$	0.0108	0.0412	0.0744	0.0033(24%)	0.0134(11%)	0.0231(8%)
$\beta = 0.001$	0.0567	0.1803	0.2754	0.019(25%)	0.0714(11%)	0.1153(7%)
$\beta = 0.0015$	0.1478	0.3682	0.4815	0.0566(25%)	0.187(9%)	0.2746(5%)
$\beta = 0.002$	0.2709	0.529	0.6257	0.1197(21%)	0.3325(7%)	0.4417(3%)

 Table 1
 Fraction of infected individuals on the contact network on day 5

The rows denominate the constants used to compute the edge infection probabilities, while the columns indicate the size of the initially infected set and the selection strategy. The values in parenthesis for the random strategy indicate standard deviation

set is small: depending on the selection of the individuals the effect on the network can be significantly different. However if the prevalence of the outbreak is high, the selection becomes less important, to the point, that it does not matter how the sources of the infection are chosen, i.e., the network "takes care" of the spreading process. We examined the expected number of infected passengers on each individual day and found that in all infection scenarios the outbreak progresses in a near-linear fashion on subsequent days, with the slope of the ascent depending on the initial conditions. There is little difference between the individual scenarios.

7 Vehicle Trip Network Analysis

The previous section explored the evolution of the outbreak across individuals in the network. However, the main objective of this study is to identify the set of vehicle trips most likely responsible for spreading infection amongst passengers (and more generally within the region) during an outbreak. We show two ways for achieving this. The first one relies solely on the simulated infection model results from the passenger contact network, while the second relies solely on the vehicle trip network, and does not require the use of an infection model.

7.1 Simulation-based Vehicle Trip Rankings

Results from the infection model defined in Section 5 provides the likelihood of infection for each node of the contact network (the individual passengers) at each point in time. Using the known assignment of the individuals to vehicle trips and the infection probability for each individual passenger, we compute an infection value for each vehicle trip by summing the probability of infection for all passengers on a given vehicle trip. Note, that while this value represents the risk of getting infected on each vehicle trip. Note, that while this value is proportional to the likelihood of infection on the vehicle trips of the transit system we rank the vehicle trips according to their infection values in a descending order and consider the fraction of the vehicle

	TOD 57 (400)	TOD 100 (000)		
	TOP 5% (400)	TOP 10% (800)	TOP 20% (1600)	
Random, $ A_0 = 10$	11.5%	20.12%	37.18%	
Random, $ A_0 = 50$	31.5%	43.5%	56%	
Random, $ A_0 = 100$	39.25%	48.75%	62.06%	
Targeted	60.5%	65.87%	73.62%	

 Table 2
 The percentage of common vehicle trips in every top 5-10-20% individual ranking in the infection scenarios

The values in parenthesis indicate the number of vehicle trips in the top 5-10-20% of the rankings. Listed are the random scenarios where the number of initially infected sources is 10,50 and 100, as well as the average across all targeted scenarios

trips on the top of the ranking. The same process is conducted for each infection scenario simulated. We refer to this ranking as the simulation-based ranking.

Our findings indicate that a significant fraction of the vehicle trips appear among the most risky trips in all infection scenarios regardless of its initial parameters. To show this, we measure the overlap between the top 5, 10 and 20% most risky vehicle trips across the individual infection scenarios. The most interesting behavior we can observe is the difference between the random scenarios, more specifically the relationship between the most likely infected vehicles and the size of the initially infected set of nodes.

Table 2 shows the fraction of vehicle trips present in every top 5-10-20% individual ranking in the infection scenarios shown in the leftmost column. These are all random scenarios with $A_0 = 10, 50$ and 100, as well as all targeted scenarios independent of the size of the initial set. There are 40 random scenarios each with fixed initial infection sizes, and 12 targeted scenarios in total. Similar to pattern of the standard deviation previously illustrated, the less infectious scenarios (small $|A_0|$) are more diverse, while the scenarios with high expected infections have a significant number of risky vehicle trips in common. This means as the number of infection sources increases, the result of the infection process converges, i.e., the network transfers the infection in such a way that the end result will be similar. This phenomenon is even more pronounced with the targeted scenarios. The underlying mechanism is very simple and well-known: the most central nodes of the network are also the most likely to be infected (Albert and Barabási 2002), and more initial sources have more opportunities to reach central nodes.

To provide an even stronger comparison, Table 3 shows the fraction of vehicle trips that are present in at least 90% of the rankings. The structure of the table is the same as Table 2. The results reinforce previous observations, i.e., the set of vehicle trips that are most likely to carry a significant number of infected passengers is robust across all tested infection scenarios.

In order to illustrate the robustness of our method towards noisy inputs, e.g., uncertainty with regards to passenger infection states, we added two levels of uniform white noise to the infection probabilities of passengers: 10 and 25%. Then we compute the infection values of the vehicle trips and identified the common trips in every top 5-10-20% individual ranking in the infection scenarios. Our results show that

	TOP 5% (400)	TOP 10% (800)	TOP 20% (1600)
Random, $ A_0 = 10$	32.25%	45.5%	59.56%
Random, $ A_0 = 50$	55.75%	66.62%	75.31%
Random, $ A_0 = 100$	62.5%	70%	79.12%
Targeted	69.75%	75.62%	82.57%

 Table 3
 The percentage of common vehicle trips in at least 90% of the top 5-10-20% individual ranking in the infection scenarios

The values in parenthesis indicate the number of vehicle trips in the top 5-10-20% of the rankings. Listed are the random scenarios where the number of initially infected sources is 10,50 and 100, as well as all targeted scenarios

introducing noise slightly decreases the overlap between the most risky vehicle trips. If the noise is small (10%) this decrease is around 1% in average, and even if the noise level is high (25%) the decrease remains below 8% in average, with the overlap becoming smaller between the more infectious scenarios. The fraction of common vehicle trips for high noise level can be seen on Table 4. Results for small noise levels are omitted because of space limitations.

7.2 Metric-based Vehicle Trip Rankings

The main contribution of this paper is to illustrate that it is possible to identify the most critical vehicle trips using only the structure of the vehicle trip network and computed network-based statistics. To do so, we computed the following network-based statistics on the vehicle trip networks:

- 1. Degree, both directed and undirected,
- 2. Node strength, both directed and undirected,
- 3. Betweenness centrality, both directed and undirected,
- 4. Eigencentrality, both directed and undirected,
- 5. PageRank values, directed by definition,
- 6. The hub/authority values of Kleinberg, directed by definition.

scenarios with high (25%) hoise rever					
	TOP 5% (400)	TOP 10% (800)	TOP 20% (1600)		
Random, $ A_0 = 10$	10.5%	18.62%	34.25%		
Random, $ A_0 = 50$	26%	35.87%	46.18%		
Random, $ A_0 = 100$	30.25%	40.25%	54.31%		
Targeted	45.5%	54.37%	64.43%		

Table 4The percentage of common vehicle trips in every top 5-10-20% individual ranking in the infectionscenarios with high (25%) noise level

The values in parenthesis indicate the number of vehicle trips in the top 5-10-20% of the rankings. Listed are the random scenarios where the number of initially infected sources is 10,50 and 100, as well as the average across all targeted scenarios

For metrics 1 to 4 both directed and undirected values were computed and compared. In addition to the network-based rankings we also computed the number of passengers that traveled on each vehicle trip resulting in the vehicle load ranking. The vehicle trip load ranking provides us with a baseline for comparison i.e. we want to improve upon these predictions with the help of the vehicle trip network.

The vehicle trips of the network were ranked according to each individual metric. We formally compare the pairwise rankings found using each network-based statistics with the simulation-based rankings using the Kendall rank correlation coefficient, defined as $\tau = \frac{n_c - n_d}{\sqrt{((n_0 - n_1)(n_0 - n_2)}}$, where $n_0 = n(n - 1)/2$, $n_1 = \sum_i t_i(t_i - 1)/2$, $n_2 = \sum_j u_j(u_j - 1)/2$, and n_c denotes the number of concordant pairs, n_d the number of discordant pairs, t_i the number of tied values in the i^{th} group of ties for the first quantity and u_j the number of tied values in the j^{th} group of ties for the second quantity. The rank correlation coefficient is used to compare rankings for each metric on both vehicle trip networks with the simulation-based rankings, for all infection scenarios, resulting in 8 * 2 * 24 = 384 different τ values. Due to space limitations, we only present the most interesting results, and provide a summary for the rest.

In Section 4, two definitions were given for the construction of the vehicle trip network. To demonstrate the difference between their usefulness in predicting the vehicle trips most likely to get infected we compare the simulation-based rankings of the previous section with rankings based on metrics computed on both vehicle trip networks. Figure 4 shows the Kendall rank correlation values between the infection and eigencentrality-based rankings for all infection scenarios for both vehicle trip networks. The correlation values are high for all scenarios further confirming the robustness of the results seen in the previous section. We can also see that the additional edges introduced by approach two clearly provide an advantage in prediction; in this specific example the improvement is 4.5% in average. The observed advantage of approach two is present for for all other metrics, therefore in the rest of the section we are only going to work with the vehicle trip network created with approach two.

We compared the network-based rankings of approach two with the infection values of the vehicle trips. The three metrics that were closest to the infection values were, in order: 1. directed eigencentrality, 2. undirected node strength, 3. vehicle trip loads. All other metrics had worse correlations than vehicle trip load, the worst being betweenness centrality.

Figure 5 shows rank correlation values between the infection rankings and vehicle load, node strength and eigencentrality. We can see, that simply considering the vehicle loads takes us close to the infection results. This is reasonable, since more frequented buses provide better means to transmit diseases among the passengers. Node strength values are in most cases marginally greater than the corresponding load correlation values. This is not surprising since node degree and strength in the vehicle trip network is associated with the number of transfer passengers from one trip to another, and vehicle trips with a lot of people on them are expected to have a greater number of transfer passengers. On the other hand, node strength only takes the number of transfer passengers into account as opposed to the number of passengers actually traveling on the trip, implying a more complex relationship. Since the closed and relatively crowded space on buses is ideal for disease spreading the total



Fig. 4 Correlation between infection values and eigencentrality for both vehicle trip networks. Targeted scenarios can be seen on (a) random ones on (b). Infection scenarios are denoted as roman numerals according to the table to the right

number of passengers traveling on a bus is proportional to the likelihood of getting infected. The actual spreading process however – the transmission of the disease from bus to bus – depends on the transfer passengers, since they are the "vectors" of the infection. These two factors explain the similarity of the values on Fig. 5.

The network metric that has the greatest correlation with the infection results was eigencentrality. Correlation values are above 0.8 for most of the scenarios indicating a close relationship. Values are typically greater with infection scenarios with greater expected infections even for the random scenarios. This is important because in the targeted scenarios, central nodes are selected as infection sources, so some form of correlation could be expected. Random ones do not have this property, yet still correlate well with the metric. Consequently, the eigencentrality ranking gives a good approximation of the infection results. The main advantage of this approach is that both the vehicle trip network and the metric can be computed quickly i.e. in a matter of seconds as opposed to the runtimes we have seen in Section 6.



Fig. 5 Correlation between infection values, node strength and eigencentrality for both vehicle trip networks. Targeted scenarios can be seen on (a) random ones on (b). Infection scenarios are denoted as roman numerals according to the table to the right

In order to illustrate the robustness of the proposed metric based analysis for the vehicle trip network, we introduced noise to the edge weights of the network to represent potential inaccuracies of passengers travel patterns. Similarly to Section 7.1 we considered two noise levels: 10 and 25%. We computed the eigencentrality metric on the noisy networks and measured the correlation between them and infection value-based rankings of the previous section. For a small noise level (10%) the decrease in correlation was below 0.3% for all infection scenarios, and below 1.5% for all scenarios in the case of a large noise level (25%). These results illustrate the highly robust nature of the proposed methods.

Figure 6 shows the map of the vehicle trips with the highest likelihood of infection in the two scenarios as well as the vehicle trips with the highest eigencentrality. The maps indicate that the vehicle trips with the highest eigencentrality highly correlate with those highly infected in the random scenario. Indeed, 6 out of the 7 transit routes in the two categories are the same. These common routes are 7, 9, 14, 17, 25, and 61 while route 94 is the one infected in random scenario but not having the highest



Fig. 6 Vehicle trips with a highest infection likelihood in targeted scenario, b highest infection likelihood in random scenario, and c highest eigencentrality

eigencentrality. Moreover, most of the vehicles trips with the highest eigencentrality appear in the set of highly infected vehicle trips in targeted attack. Indeed, all 7 transit routes with the highest eigencentrality trips appear in the 9 routes with high infection in the targeted scenario. These common routes are 7, 9, 11, 14, 17, 25, and 61, while routes 4 and 94 are those infected in targeted scenario and not having the highest eigencentrality.

8 Conclusions and Future Works

In this paper, we presented a novel method to identify the components of a public transit system which are most likely to transport infected passengers, and therefore play a vital role in the continued spread of infection during an outbreak. Based on the contact network of passengers traveling on public transportation system, we defined a new network structure, the vehicle trip network, which is capable of representing

the system in a compact way. Re-defining the passenger movements using this novel network structure provides a means to efficiently identify the critical components of the network which should be targeted for surveillance and control, in efforts to mitigate an ongoing outbreak. We demonstrated the usefulness of the new structure in predicting the most critical vehicle trips and routes in the network by first running a variety of infection results on the contact network, and then showing that it is possible to identify the most risky trips during an infection by computing simple network metrics on the trip network.

Based on the results from a case study using the public transit network from Twin Cities, MN, the set of transit vehicle trips at highest risk of infection is robust to the initial conditions of the outbreak, and it is possible to identify of the vast majority these trips by simply computing the eigencentrality metric on the vehicle trip network. Furthermore, the methods are illustrated to be robust to two types of data uncertainty, those being passenger infection levels and travel patterns of the passengers. The vehicle trips identified by the model to be at highest risk would be optimal locations to implement vehicle surveillance, for example in the forms of temperature gauges such as those recently installed on buses in Hong Kong. Identifying the critical spreading components of the transit system can aid public health authorities in the optimal allocation of surveillance resources, and is essential in mitigating future outbreaks.

A natural continuation of this work is to design and implement an infection model on the vehicle trip network. There are two challenges to this approach. The vehicle trip network in its current form represents only the transfer passengers of the transportation network, but an infection process involves the bulk passengers as well. To solve this, the infection model has to be modified to "model" the behavior of both types of passengers during the outbreak process. The second challenge is in making the connection between contact and vehicle trip networks. Since the links of the vehicle trip network correspond to sets of transfer passengers on the contact network, it is natural to assume, that the edge infection probabilities of the vehicle trip network are related to the infection probabilities of the corresponding transfer passengers. The exact nature of this relation, however, has to be investigated in detail.

Acknowledgements We thank the National Health and Medical Research Council (NHMRC) for funding, project grant (No. APP1082524). The contents of the published material are solely the responsibility of the Administering Institution, a Participating Institution or individual authors and do not reflect the views of the NHMRC.

References

Albert R, Barabási A (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47

- Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford
- Balcan D (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA 106:21,484–21,489

Bóta A, Krész M, Pluhár A (2013) Approximations of the generalized cascade model. Acta Cybern 21(1):37-51

Bóta A, Gardner L, Khani A (2017) Modeling the spread of infection in public transit networks: a decisionsupport tool for outbreak planning and control. In: Transportation research board 96th annual meeting

Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462-465

- Cahill E, Crandall R, Rude L, Sullivan A (2005) Space-time inuenza model with demographic, mobility, and vaccine parameters. In: Proceedings of 5th annual Hawaii international conference of mathematics statistics and related fields
- Candia J, González MC, Wang P, Schoenharl T, Madey G, Barabási A (2008) Uncovering individual and collective human dynamics from mobile phone records. J Phys A Math Theor 41(224015):11
- Carley K, Fridsma D, Casman E, Yahja A, Altman N, Chen L, Kaminsky B, Nave D (2006) Biowar: scalable agent-based model of bioattacks. IEEE Trans Syst Man Cybern Part A Syst Hum 36(2): 252–265
- Cattuto C (2010) Dynamics of person-to-person interactions from distributed rfid sensor networks. PloS One 5:e11,596
- Chen N, Gardner L, Rey D (2016) A bi-level optimization model for the development of real-time strategies to minimize epidemic spreading risk in air traffic networks. Transp Res Rec: J Transp Res Board No 2569
- Christakis NA, Fowler JH (2010) Social network sensors for early detection of contagious outbreaks. PLoS One 5:e12,948
- Coleman J, Menzel H, Katz E (1996) Medical innovations: a diffusion study. Bobbs Merrill, New York
- De Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: The privacy bounds of human mobility. Sci Rep 3:1376
- Dibble C, Feldman PG (2004) The geograph 3d computational laboratory: network and terrain landscapes for repast. J Artif Soc Soc Simul 7(1)
- Dunham J (2005) An agent-based spatially explicit epidemiological model in mason. J Artif Soc Socx Simul 9(1):3
- Ekici A, Keskinocak P, Swann J (2008) Pandemic influenza response. In: Winter simulation conference, pp 1592–1600
- Epstein JM, Cummings DAT, Chakravarty S, Singa RM, Burke DS (2002) Toward a containment strategy for smallpox bioterror: an individual-based computational approach. Brook Inst Press 2004:55
- Erath A, Löchl M, Axhausen KW (2009) Graph-theoretical analysis of the Swiss road and railway networks over time. Netw Spat Econ 9(3):379–400
- Eubank S (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429:180-184
- Fajardo D, Gardner L (2013) Inferring contagion patterns in social contact networks with limited infection data. Netw Spat Econ 13(4):399–426
- Ferguson NM (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437:209–214
- Funk S, Salathé M, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 7:1247–1256
- Galvani AP, May RM (2005) Epidemiology: dimensions of superspreading. Nature 438:293-295
- Gardner L, Sarkar S (2013) A global airport-based risk model for the spread of dengue infection via the air transport network. PLoS One 8(8):e72,129. doi:10.1371/journal.pone.0072129
- Gardner L, Fajardo D, Waller S (2012) Inferring infection-spreading links in an air traffic network. Transp Res Rec: J Transp Res Board 2300:13–21
- Gardner L, Fajardo D, Waller S, Wang O, Sarkar S (2012) A predictive spatial model to quantify the risk of air-travel-associated dengue importation into the United States and Europe. J Trop Med 2012:103,679. doi:10.1155/2012/103679
- Gardner L, Fajardo D, Waller S (2014) Inferring contagion patterns in social contact networks using a maximum likelihood approach. Nat Hazards Rev 15(3)
- Gastner M, Newman M (2006) The spatial structure of networks. Eur Phys J B 49(2):247-252
- Germann TC, Kadau K, Longini I, Macken CA (2006) Mitigation strategies for pandemic inuenza in the United States. Proc Natl Acad Sci 103(15):5935–5940
- Gilbert MT (2007) The emergence of hiv/aids in the americas and beyond. Proc Natl Acad Sci USA 104:18,566–18,570
- González M, Lind P, Herrmann H (2006) System of mobile agents to model social networks. Phys Rev Lett 96(8):088,702

- González MC, Hidalgo CA, Barabási AL (2008) Understanding individual human mobility patterns. Nature 453:779–782
- Hasan S, Ukkusuri S (2011) A contagion model for understanding the propagation of hurricane warning information. Transp Res Part B 45(10):1590–1605
- Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith J, Woolhouse MEJ (2003) The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proc R Soc B 270:121–127
- Hoogendoorn S, Bovy P (2005) Pedestrian travel behavior modeling. Netw Spat Econ 5(2):193-216
- Huerta R, Tsimring LS (2002) Contact tracing and epidemics control in social networks. Phys Rev E Stat Nonlin Soft Matter Phys 66(056):115
- Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci USA 101(42):15,124–15,129
- Illenberger J, Nagel K, Flötteröd G (2012) The role of spatial interaction in social networks. Netw Spat Econ 13(3):1–28
- Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence though a social network. In: Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146
- Khani A, Hickman M, Noh H (2015) Trip-based path algorithms using the transit network hierarchy. Netw Spat Econ 15(3):635–653
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. Nat Phys 6:888–893
- Kuiken C, Thakallapalli R, Eskild A, De Ronde A (2000) Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. Am J Epidemiol 152:814–822
- Lam WK, Huang H (2003) Combined activity/travel choice models: time-dependent and dynamic versions. Netw Spat Econ 3(3):323–347
- Meyers L, Pourbohloul B, Newman MEJ, Skowronski D, Brunham R (2005) Network theory and sars: predicting outbreak diversity. J Theor Biol 232:71–81
- Murray J (2002) Mathematical biology, 3rd edn. Springer, New York
- Nassir N, Khani A, Hickman M, Noh H (2012) An intermodal optimal multi-destination tour algorithm with dynamic travel times. Transp Res Rec: J Transp Res Board 2283:57–66
- Pendyala R, Kondhuri K, Chiu YC, Hickman M, Noh H, Waddell P, Wang L, You D, Gardner B (2012) Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation. Transp Res Rec: J Transp Res Board 2203:19–27
- Ramadurai G, Ukkusuri S (2010) Dynamic user equilibrium model for combined activity-travel choices using activity-travel supernetwork representation. Netw Spat Econ 10(2):273–292
- Rey D, Gardner L, Waller S (2016) Finding outbreak trees in networks with limited information. Netw Spat Econ 16(2):687–721
- Roche B, Drake J, Rohani P (2011) An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. BMC Bioinforma 12(1):87
- Roorda M, Carrasco J, Miller E (2009) An integrated model of vehicle transactions, activity scheduling and mode choice. Transp Res Part B 43(2):217–229
- Rvachev L, Longini I (1985) A mathematical model for the global spread of influenza. Math Biosci 75:3–22
- Salathé M (2010) A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci USA 107:22,020–22,025
- Schintler L, Kulkarni R, Gorman S, Stough R (2007) Using raster-based gis and graph theory to analyze complex networks. Netw Spat Econ 7(4):301–313
- Small M, Tse C (2005) Small world and scale free model of transmission of sars. Int J Bifurcations Chaos Appl Sci Eng 15(1745)
- Song C, Qu Z, Blumm N, Barabási AL (2010) Limits of predictability in human mobility. Science 327:1018–1021
- Stehlé J (2011) Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. BMC Med 9:87
- Sun L, Axhausen KW, Lee DH, Huang X (2013) Understanding metropolitan patterns of daily encounters. Proc Natl Acad Sci USA 110:13,774–13,779
- Troko J, Myles P, Gibson J, Hashim A, Enstone J, Kingdon S, Packham C, Amin S, Hayward A, Van-Tam JN (2011) Is public transport a risk factor for acute respiratory infection? BMC Infect Dis 11(1):16

- Wang P, Gonzlez MC, Hidalgo CA, Barabsi AL (2009) Understanding the spreading patterns of mobile phone viruses. Science 324:1071–1076
- Wesolowski A, Buckee C, Bengtsson L, Wetter E, Lu X, Tatem A (2014) Commentary: containing the ebola outbreak–the potential and challenge of mobile network data. PLOS currents outbreaks
- Wu J, Xu F, Zhou W, Feikin D, Lin C, He X, Zhu Z, Liang W, Chin D, Schuchat A (2004) Risk factors for sars among persons without known contact with sars patients, Beijing, China. Emerg Infect Dis J-CDC 10(2):210–216