

Modelling Spreading Process Induced by Agent Mobility in Complex Networks

Wei Koong Chai, *Member, IEEE*

Abstract—Most conventional epidemic models assume contact-based contagion process. We depart from this assumption and study epidemic spreading process in networks caused by agents acting as carrier of infection. These agents traverse from origins to destinations following specific paths in a network and in the process, infecting the sites they travel across. We focus our work on the *Susceptible-Infected-Removed* (SIR) epidemic model and use continuous-time Markov chain analysis to model the impact of such agent mobility induced contagion mechanics by taking into account the state transitions of each node individually, as oppose to most conventional epidemic approaches which usually consider the mean aggregated behavior of all nodes. Our approach makes one mean field approximation to reduce complexity from exponential to polynomial. We study both network-wide properties such as epidemic threshold as well as individual node vulnerability under such agent assisted infection spreading process. Furthermore, we provide a first order approximation on the agents' vulnerability since infection is bi-directional. We compare our analysis of spreading process induced by agent mobility against contact-based epidemic model via a case study on London Underground network, the second busiest metro system in Europe, with real dataset recording commuters' activities in the system. We highlight the key differences in the spreading patterns between the contact-based vs. agent assisted spreading models. Specifically, we show that our model predicts greater spreading radius than conventional contact-based model due to agents' movements. Another interesting finding is that, in contrast to contact-based model where nodes located more centrally in a network are proportionally more prone to infection, our model shows no such strict correlation as in our model, nodes may not be highly susceptible even located at the heart of the network and vice versa.

Index Terms—Epidemic, Complex networks, carrier agent, mobility.



1 INTRODUCTION

SPREADING phenomena occur in various networks, including networks that are naturally formed (*e.g.*, hub protein, brain structure, *etc.*) and human engineered ones (*e.g.*, communication networks, power/utility grids, transportation systems, *etc.*). Diseases spread through populations via different pathogens [1] [2] or proteins [3] as well as digital diseases infecting various devices in the form of computer virus, spyware and malware [4] [5]. In (on-line) social networks, gossips, news and rumors propagate from one friend to another. For instance, viral marketing has recently been used to propagate information from one "friend" to another in social media [6]. Failures in man-made critical infrastructures are also found to cascade from one network to another due to increased inter-dependency between the originally isolated networks [7] [8] [9]. These spreading phenomena, and many others, can either be desirable (*e.g.*, alerts/warnings, news, information *etc.*) which may be promoted or harmful (*e.g.*, virus, diseases, riots, *etc.*) which should be contained¹. Such phenomena have been the subject of study in diverse fields such as biology, sociology, physics, computer science, *etc.* In most of these studies, contagion process are assumed to be relying on contacts between nodes (possibly in different modes – *e.g.*, physical meetings, electronic communications, *etc.*). The dif-

fusion pattern is usually omni-directional. The spreading of information is then solely dependent on *who meets with who*. This then can be accounted by identifying the neighbours of each individual in the network which is equivalent to node degree. While classic results (*e.g.*, [4] [10]) mostly associate a "contagion" process with the structural properties of the network (*i.e.*, the first and second moments of the network degree distribution), new models have shown that epidemic behaviour is more accurately determined by the spectral properties of the network [11] [12] [13], thus bringing spectral graph theory into the analysis.

In this work, we focus on *simple* spreading processes² which do not require multiple and/or different instances of exposure before a state transition takes place (*e.g.*, a healthy node becoming infected). Specifically, we are interested in understanding simple contagion process following the *Susceptible-Infected-Removed* (SIR)³ epidemic model [15] that is induced by carrier agents that move along certain prescribed paths within the network. One example of this class of diffusion phenomena is on rail [16] and air [17] transportation systems where human beings (and in some cases, pets) become the carrier agents transporting the contagion. In fact, it has already been found that human mobility follows highly regular routine patterns (*e.g.*, route between home and workplace) [18] [19]. The advent of smartphones also open a whole new range of avenues for computer viruses and malware to spread. Infected user

• W. K. Chai is with the Department of Computing and Informatics, Bournemouth University, Dorset, United Kingdom, BH12 5BB. E-mail: wchai@bournemouth.ac.uk

1. For the rest of this paper, we will use the word "information" as the generic term referring to any object that is being disseminated in the network (*e.g.*, tweets in online social network, multimedia content in the Internet, malware in computer networks, contagious diseases in a population, *etc.*).

2. As oppose to *complex* spreading processes [14] such as technology adoption, influence spreading, *etc.*

3. The *R* state in the *SIR* model is also often referred to as the "recovered" state. For the purpose of this paper, we treat them as the same state.

devices move along the paths the owners are traveling. The virus/malware then actively attempt to replicate itself to nearby devices through wireless communication technologies such as Wi-Fi or Bluetooth. For instance, Cabir⁴ and CommWarrior⁵ are two worms discovered in mid-2000 that replicate themselves to other devices via Bluetooth. Recent studies such as [20] and [21] have predicted serious threat to mobile communications especially within a public transportation network. In the emerging information-centric networking (ICN) paradigm [22], information or content is being cached in routers equipped with cache stores along the paths they traverse [23] [24]. Computer viruses and malware may also spread via piggybacking email chains. An infected nodes may later become immune or simply choose not to further disseminate the information gained from the agents.

In this paper, we advance the current contact-based epidemic modelling approach by considering a different infection spreading process that is carried by infectious agents. We explicitly model the role played by such agents on the spreading dynamics. Following such considerations, the susceptible nodes in the network will possess different probability of getting infected as the amount of interactions in the network system governs that infection potential. A node having higher volume of infectious agents destined to or traversing it will have higher chances of contracting the infection. The node degrees will no longer have direct influence to the contagion process since the mere existence of a link between two nodes does not automatically results in infection; only when interactions happen between the two nodes (which is made possible by the presence of the link) will cause infection to be passed.

Our main contributions are as follows. We contribute to the theoretical development of epidemic theory and model the spreading dynamics caused by infectious agents traversing the network (*i.e.*, agent mobility) using graph theoretic approach. Our model break from the conventional assumption of the contact-based infection process and account for the prescribed routes the agents take between two network nodes. In our model, a node may still be susceptible to infection even when all its immediate neighbors are uninfected. Spreading is now dictated by the amount of activity and mobility pattern of the agents. Our approach takes the analytical framework studied in [12] [25] as the starting point. In our model, we create an infection characterization matrix that allows us to draw insights into the behavior of the epidemic. We find the epidemic threshold of such contagion process as well as the upper and lower bounds of this threshold for a given infection characterization matrix. Several interesting conclusions are drawn. Unlike contact-based epidemic, spreading due to agents' mobility affects more nodes across the network. Moreover, another finding relates to the location of nodes in the network whereby nodes that lie near the edges of the network are not necessarily less susceptible to infection. This is in direct contrast with (weighted) contact-based epidemic models which consistently predict high infection probabilities for "central" nodes such as hubs with high number of neighbors.

4. <https://www.f-secure.com/v-descs/cabir.shtml>

5. <https://www.f-secure.com/v-descs/commwarrior.shtml>

TABLE 1
Notations

Symbol	Descriptions
A	Adjacency matrix representing the topology
V	The set of nodes (or vertices) in the network
E	The set of links (or edges) in the network
N	Number of nodes in the network
L	Number of links in the network
β	Infection probability
μ	Remove/recovering rate
$s_n(t)$	Probability of node n in the healthy state
$i_n(t)$	Probability of node n in the infected state
$r_n(t)$	Probability of node n in the removed state
$\rho(t)$	Fraction of infected nodes in the network at time t
Q_n	Infinitesimal generator of node n 's continuous Markov chain
$q_{i,j;n}$	The (i, j) element of Q_n
τ_c	Epidemic threshold
$\lambda_{n,m}$	Number of agents originated from v_m and destined to v_n
Λ	Matrix encoding the agent traffic, $\lambda_{n,m}$, for all node pairs
$P_{n,m}$	The set of all nodes involved in the path between v_m and v_n inclusive.
$\mathbf{1}_x$	Indicator function; Equals 1 when $x = \text{true}$ and 0 otherwise.
C	Infection characterization matrix
p	Probability of a link connects a pair of nodes (for generating random graphs)
w	The number of links to attach at each iteration (for generating scale-free graphs)
γ_1	Largest eigenvalue of matrix C

The rest of the paper is organized as follows. In Section 2, we review the fundamentals of epidemic models and the related work to modelling epidemics, leading to our specific epidemic spreading model. We detail the exact spreading mechanics that we consider in this work in Section 3. Based on the specified spreading mechanics, we develop the modelling framework next in Section 4 that is of polynomial complexity and provide Monte-carlo simulation results to validate our model. We then discuss in Section 5 the spreading process induced by agents traversing the network from three perspectives, namely from the point of view of the network, individual nodes and the agents respectively. Following this, we present in Section 6 our analysis of a case study on a hypothetical epidemic spread via commuters (as the agents) traveling in the London Underground network system and highlight the main findings of our model compared against conventional contact-based epidemic contagion process. Finally, we conclude our study in Section 7. Table 1 summarizes the notations used in this paper.

2 BACKGROUND, BASICS AND RELATED WORK

In recent years, epidemic theory has found applications in various different fields, covering virus/disease spreading (both biological and digital ones) (*e.g.*, [1] [5]) and corresponding immunization strategies (*e.g.*, [26] [27] [28] [29]), information dissemination in (online) social networks (*e.g.*, [30]), communication protocol design (*e.g.*, [31] [32] [33]) and cascading failure prediction/protection (*e.g.*, [34]) as well as in more general contexts, analysis on stability of spreading processes over time-varying networks (*e.g.*, [35]) and iden-

tification of influential seeds/spreaders in networks (e.g., [36]).

Initial theoretical treatment on epidemic can be traced back to Bernoulli's work on smallpox in 1760 [37]. Early epidemic models are deterministic which are especially suitable for predicting epidemic for large populations [10]. Epidemic spreading is inherently a random process and thus, stochastic models were later proposed with [15] being one of the earliest. The research in the literature focused on theoretically expressing the process of how a susceptible individual contract the epidemic via contacts with infected ones. One well-known initial effort is to invoke the so-called homogeneous mixing assumption [15] whereby all individuals in the population are assumed to have equal probability of meeting each other.

Epidemic theory has now been considered as part of network science [38]. Conventional epidemic models are usually described via a set of ordinary differential equations (ODEs) whereby infection is passed between individuals by *contact* – an individual “meeting” another. Commonly, the population is first segregated into groups based on the state of the individuals. Each individual can and must assume one of the possible distinct states (or stages) of the epidemic. Some common possible states are susceptible (S), infectious (I), removed (R), exposed (E), alert (A), etc.

Since we are considering *SIR* epidemic model here, each individual in the network must assume one and only one of the following states at any time:

- Susceptible (S) – Individuals who are free from but prone to infection.
- Infected (I) – Individuals who are infected. They are also assumed to be infectious.
- Removed or Recovered (R) – Individuals who are immune from infection. They are also assumed to be not infectious.

Then, a state transition rate (usually a constant) that is independent of the contacts amongst individuals is defined. The transition rate can then be written as $dX(t)/dt = -\alpha X(t)$ where $X(t)$ generically represents a group of individuals being in a possible state X at time t and α is the constant inverse of the average period an individual spent in this group or the per capita transition rate. Under the homogeneous mixing assumption, the infection rate then simply depends on the size of the susceptible and infected groups, usually given by $\alpha S(t)I(t)$. With this assumption, the heterogeneities of individuals are neglected. Following such complete random mixing context, the (now classic) homogeneous ODE for an *SIR* model can be written as follows [15]:

$$\frac{ds}{dt} = -\beta si \quad (1)$$

$$\frac{di}{dt} = \beta si - \mu i \quad (2)$$

$$\frac{dr}{dt} = \mu i \quad (3)$$

where β represents the infection transmission rate, μ is the removal/recovery rate and finally, s , i and r denote the fraction of population in susceptible, infected and removed state respectively.

However, real world phenomena are mostly filled with heterogeneities. Appreciating this feature, the role of heterogeneities have then been considered. For instance, [1] considered age, genetic and social heterogeneities in the host population for sexually transmitted diseases. Heterogeneity with respect to network structures were also considered in [39] [40] [41] [42]. Individual nodes are segregated into classes of equivalence. Nodes with the same number of immediate neighbors (node degrees) are considered to be in the same class and are assumed to be statistically equivalent. Often, the well-known *configuration model* [43] is used to illustrate the accuracy of the proposed model since with this model, specific degree distribution can be specified by giving the normalized probabilities that a random node has certain degree.

Several pair-wise approximation models were proposed to model epidemic spread with better accuracy at the expense of being more mathematically involved than previous models employing mean-field approximations [44] (Chapter 7.8). In [45], a basic pair-wise approximation model is proposed; assuming uniform contact distribution and ignoring clustering effect of individuals. The pair-wise approximation approach was further developed with network heterogeneities taken into account in [46] [47]. Furthermore, in [48], the authors described a pair-approximation mathematical model for worm propagation that considered the network features such as degree distribution and clustering coefficient. This approach, however, come with the cost of higher computational complexity (i.e., the model requires consideration of $\binom{N}{2}$ number of pairs).

These above works in the literature usually employ some approximations techniques to reduce computational complexity since the complexity to compute the exact solution of the epidemic spread grows exponentially with the number of nodes in the network [12]. The authors in [12] further developed an N -intertwined modelling approach that reduces the complexity of the exact solution for *SIS* epidemic spreading from exponential to polynomial. This approach is extended to model the *SIR* epidemic in [49]. Our work presented here diverges from the contact-based assumption used in these earlier works. We consider infection being carried by an infected agent moving from one node to another along certain path⁶.

The abovementioned works form a solid foundation to the modelling of epidemic and allow us to gain significant understanding into the characteristics of epidemic dynamics. As elucidated in Section 1, in many real world scenarios, pure contact-based contagion assumption may not be sufficient as in our considered case here where infection may be carried across certain path due to agents' movements. In this direction, a closely related class of epidemic models is the one that considers spread of infection by individual movements between distinct sub-populations or locations [44]. Among the different spatial models, the concept of metapopulation is one of the most versatile and used in the literature. For instance, contact-based epidemic dynamics was studied in metapopulation models in [50] by assuming

6. Readers may find it helpful to map the distinction between contact-based vs. mobility-based contagion viewpoints with *Lagrangian* vs. *Eulerian* viewpoints in the classical field theory

individuals having random Markovian movements in the subpopulation without memory of their origin, in [51] by assuming bi-directional mobility of individuals between their unique base location and a small subset of frequently visited locations and in [52] by considering recurrent mobility patterns. The spreading behaviour based on store-and-forward paradigm (as used in delay-tolerant networks) falls under such models where infection is still assumed to be contact-based but with added considerations of node movements. Furthermore, the role of agents in spreading infection is considered in [53] whereby the number of individuals traveling between subpopulations required for macroscopic infection spread is analyzed. The effect of short- and long-range traveling patterns on spatial infection spreading was investigated in [54] while the effect of restricted mobility (*i.e.*, travel restrictions) in controlling infection spreading was studied in [55]. In our work, we do not assume random / recurrent / restricted mobility pattern as above. Instead, we encode the infectious agents' traveling paths in an infection characterization matrix. In view of the need for large-scale simulations in epidemic studies, a comparison between two popular methodologies (*i.e.*, stochastic agent-based models and structured metapopulation models) were made in [59], showing good simulation agreement between the two modeling approaches. In the area of communication networks, [56] modelled epidemic dynamics by considering data packets as the infection agents traveling between random source-destination pairs. Taking [56] as the modelling basis, [57] and [58] have studied the question on how to control epidemic spreading by link removals and via different routing schemes respectively. These works assumed stochastic interactions between node pairs. We depart from such assumption and model explicitly the interactions. In our work, we differ methodologically by employing graph theoretic approach instead to model the contagion process, taking into account *exogenous* individual arrivals with each individual traveling across a path.

3 SPREADING MECHANICS VIA CARRIER AGENTS

In this section, we describe the spreading mechanics facilitated by carrier agents' mobility. Specifically, we consider an agent in the network can assume one of the two states at any one time (*i.e.*, either susceptible or infected). Once infected, the agent stays being infected until it reaches its destination. As such, the agents follow the *SI* epidemic model. To simplify our model, we assume Poisson agent arrival process with no routing delay.

We use Fig. 1 to illustrate the role of agents in spreading infection. When there is no influence of agents (Fig. 1(a)), susceptible and removed individuals will not change their state. However, an infected individual may change to the removed state with a removal rate, μ . This transition is assumed to be a Poisson process independent of the agent traffic. As per convention in the literature, and without loss of generality, we use $\mu = 1.0$ for the rest of the paper.

An agent is infected and stays infectious if it originates from an infected node (Fig. 1(b)). Conversely, if originated from a susceptible or removed node, the agent assumes the susceptible state. An infected agent infects a susceptible node with transmission rate β . In our model, an agent may

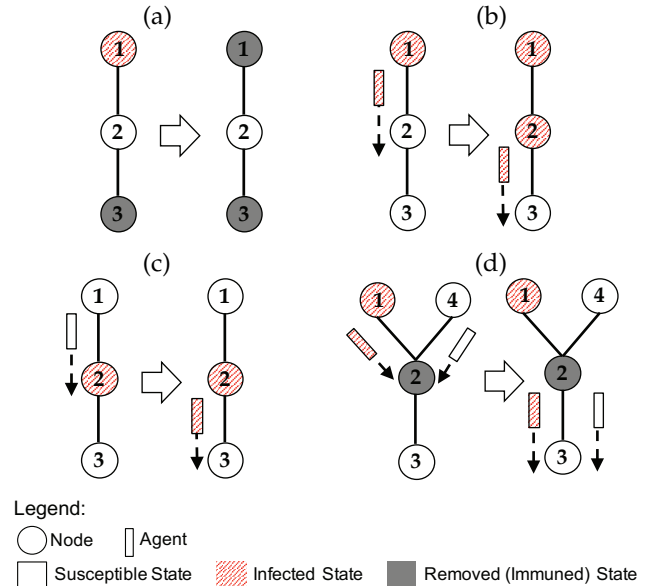


Fig. 1. (Color online) Illustration of possible ways for a state transition to take place. (a) No agent carrying infection will not cause new infections. Susceptible and removed nodes do not change state. Infected nodes may become immuned and thus assume the removed state. (b) Agents originated from an infected node are also infected. Infected agents are infectious. (c) Infected node can infect agents traversing it. (d) Immuned nodes do not change the state of the agent and vice versa, agents do not change the state of immuned nodes.

infect multiple nodes along the path in one transmission. A susceptible agent is infected once it reaches any infected node (Fig. 1(c)). As such, the longer the path, the higher the probability the agent traversing this path will be infected. Finally, removed and susceptible nodes do not change agents' state (Fig. 1(d)).

4 MODEL DEVELOPMENT

We consider an underlying undirected network, $G(V, E)$ with $V = v_1, \dots, v_N$ nodes and $E = e_1, \dots, e_L$ links where $N = |V|$ and $L = |E|$. G can be represented by A , the $N \times N$ symmetric adjacency matrix, with $a_{n,m} = 1$ if there exists a link between node n and m and 0 otherwise. In this work, one may think of the nodes as individuals, devices or locations depending on the considered scenario while links as the channel allowing direct access between two nodes.

For simplicity, we use the term *agents* as the universal infectious entity that passes the infection by traversing different nodes via the connecting links. We model the spreading process that is based on the paths of agent flows. The "infection" can be transmitted by different kinds of agents depending on the specific application context (*e.g.*, content chunk in the case of in-network caching, software patch in the case of computer virus immunization, tweets in the context of gossip spreading in online social networks, *etc.*). The exact mechanics of the infection process are described in Section 3.

Let $s_n(t)$, $i_n(t)$ and $r_n(t)$ denote the probability of node n being in the healthy, infected and removed state at time, t , respectively. Following the fact that each node must be in one of the three states at any time t , then

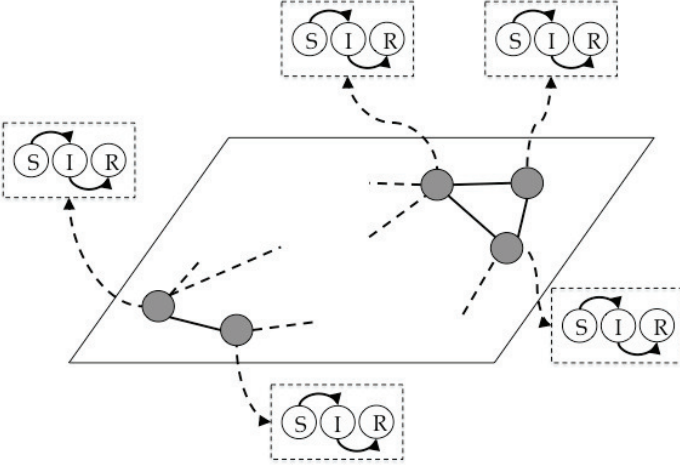


Fig. 2. Each node in the network can be in either one of the three states (i.e., S, I, R) at one time. The combinations of all possible states is 3^N .

$$s_n(t) + i_n(t) + r_n(t) = 1 \quad (4)$$

and

$$\frac{ds_n(t)}{dt} + \frac{di_n(t)}{dt} + \frac{dr_n(t)}{dt} = 0. \quad (5)$$

Now, if we directly apply Markov theory to the entire network, then the complexity of the solution is exponential (i.e., $O(3^N)$) since we need to consider all the possible combinations of states for each and every node and in this case, the infinitesimal generator of the system, $Q(t)$, has the dimension of $3^N \times 3^N$.

Instead, we approach the problem by considering each node individually (see Fig. 2) and applying Markov theory to obtain N copies of infinitesimal generator $Q_n(t)$ of the three-state continuous Markov chain for each node separately as follows:

$$Q_n(t) = \begin{bmatrix} -(q_{1,2;n} + q_{1,3;n}) & q_{1,2;n} & q_{1,3;n} \\ q_{2,1;n} & -(q_{2,1;n} + q_{2,3;n}) & q_{2,3;n} \\ q_{3,1;n} & q_{3,2;n} & -(q_{3,1;n} + q_{3,2;n}) \end{bmatrix} \\ = \begin{bmatrix} -q_{1,2;n} & q_{1,2;n} & 0 \\ 0 & -q_{2,3;n} & q_{2,3;n} \\ 0 & 0 & 0 \end{bmatrix} \quad (6)$$

The remove rate is a Poisson process and is not dependent on agents' activity within the network. Hence, $q_{2,3;n} = \mu$. In contrast, $q_{1,2;n}$ is dependent on the states of other nodes within the network which are further dependent on the agents traversing the nodes. To account for this dependency, $q_{1,2;n}$ must be conditioned with all possible combinations of states for all nodes, resulting back to the exact Markov chain solution which, unfortunately, is of exponential complexity.

We now proceed following the approach adopted in [12] [60] by applying a mean field approximation to the random variable, $q_{1,2;n}$ and use its expected rate instead (see Fig. 3). This removes its random nature as well as reduce the complexity of the solution to polynomial $O(N)$.

The effective infinitesimal generator can now be written as:

$$\overline{Q_n(t)} = \begin{bmatrix} -E[q_{1,2;n}] & E[q_{1,2;n}] & 0 \\ 0 & -\mu & \mu \\ 0 & 0 & 0 \end{bmatrix} \quad (7)$$

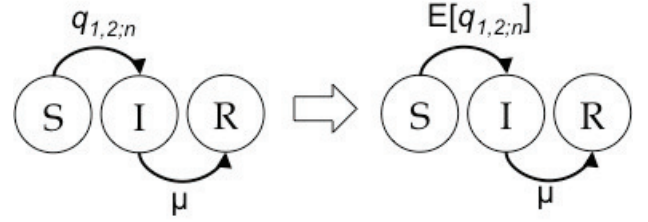


Fig. 3. Applying the mean field approximation reduces analysis complexity of exact solution from exponential $O(3^N)$ to polynomial $O(N)$.

From Eq. 7, $E[q_{1,2;n}]$ models the effective infection rate of v_n . In classic contact-based epidemic models, this is basically a function of the infection rate, β and the states of v_n 's neighbours which in turn relates to the adjacency matrix, A . Node v_n is not immediately susceptible to infection until at least one of its neighbours is infected. In our case, this is not true since the infection no longer depends on neighbours' states but rather the state of the agents traversing the node. As such, we need to model the probability of agents crossing or destined to the node of interest are carrying infection.

Let the expected volume of agents originated from v_m and destined to v_n be $\lambda_{n,m}; \forall n, m \in V; n \neq m$. As such, we have the following:

$$\forall n \in V : \lambda_{n,n} = 0. \quad (8)$$

In matrix form, Λ has the trace, $tr(\Lambda) = \sum_{n=1}^N \lambda_{n,n} = 0$ (i.e., no self-traffic). In real world scenario, the matrix Λ is sometimes known as the traffic matrix [61] that denotes the amount of expected traffic between node pairs. It can often be estimated/predicted [62] or historical data may be used (e.g., statistics collected over time such as the case with our case study in Section 6).

Since we consider a connected network, there exists at least one path between any node pair. Further, we assume there exists agents traveling between all node pairs and all agents travelling between two nodes follow the same paths at all times. We then have a total of $N(N-1)$ distinct paths.

Using the above information, we construct the infection characterization matrix, $C = (c_{n,m})$, which is an irreducible non-negative square matrix. The element $c_{n,m}$ represents the total agents flowing through or destined to node v_n originated from node v_m and can be written as follows:

$$c_{n,m} = \sum_{k=1}^N \lambda_{k,m} \mathbf{1}_{n \in P_{k,m}} \quad (9)$$

where the indicator function is defined as:

$$\mathbf{1}_{n \in P_{k,m}} := \begin{cases} 1; & \text{if } n \in P_{k,m}, \\ 0; & \text{if } n \notin P_{k,m}. \end{cases} \quad (10)$$

where $P_{k,m}$ is the set of all nodes involved in the path between nodes v_m and v_k (source and destination nodes inclusive). The membership of the set $P_{k,m}$ depends on the specific application scenario. For instance, for communication networks, this is determined by the routing protocol employed by the network operator whereas for transportation scenarios, this is dependent on the planned flight/train routes. For an illustrative example on deriving matrix C

based on the network topology (*i.e.*, A) and agent mobility (described via Λ and $P_{n,m}$).

Using Eq. 9, we can write the following:

$$E[q_{1,2;n}] = \beta \sum_{m=1}^N i_m(t) c_{n,m} \quad (11)$$

Applying the Markov differential equation (see [63] (10.11) on p. 208), we obtain the following system of non-linear differential equations,

$$\frac{ds_n(t)}{dt} = -s_n(t)\beta \sum_{m=1}^N i_m(t) c_{n,m} \quad (12)$$

$$\frac{di_n(t)}{dt} = s_n(t)\beta \sum_{m=1}^N i_m(t) c_{n,m} - \mu i_n(t) \quad (13)$$

$$\frac{dr_n(t)}{dt} = \mu i_n(t) \quad (14)$$

The epidemic process is governed by the set of differential Eqs. 12, 13 and 14. Specifically, from Eq. 13, we see two opposite processes: when node n is healthy with probability $s_n(t)$, it is being infected by total infectious agents traversing or destined to it, $\sum_{m=1}^N i_m(t) c_{n,m}$, with rate β and when node n is infected with probability $i_n(t)$, it is removed at rate μ .

This system of equations written in its matrix form is as follows:

$$\frac{dS(t)}{dt} = -S(t)\beta CI(t) \quad (15)$$

$$\frac{dI(t)}{dt} = S(t)\beta CI(t) - \mu I(t) \quad (16)$$

$$\frac{dR(t)}{dt} = \mu I(t) \quad (17)$$

where $S(t) = [s_1(t) \ s_2(t) \ \dots \ s_N(t)]^T$, $I(t) = [i_1(t) \ i_2(t) \ \dots \ i_N(t)]^T$ and $R(t) = [r_1(t) \ r_2(t) \ \dots \ r_N(t)]^T$.

We can now solve the above system of differential equations to obtain the instantaneous evolution of the population for the three distinct states. Furthermore, using Eqs. 4 and 5, we can reduce the problem from solving $3 \times N$ simultaneous differential equations to $2 \times N$ equations.

We show the agreement of our analytical framework against results from Monte-Carlo simulations for both pure random graphs such as Erdős-Rényi (ER) graph model [64], which has binomial degree distribution, and for scale-free (SF) graphs [65], which have power-law degree distribution as the literature has shown evidence that the two types of graph models exhibit different epidemic behaviours. As mentioned before, our approach takes individual node into account separately and as such, for validation purposes, we do not have to rely on network topologies generated with specific degree distributions such as using the *configuration model* [43].

For our simulations, we generate a sample set of graphs for each graph model above. For ER graphs, for a given N , a link randomly connects a pair of nodes with probability p independent of other links. Unless otherwise specified, we use $p = 2 \times p_c = 2 \times \ln(N)/N$ which is two times

the sharp threshold for connectedness to ensure graphs are fully connected while at the same time having sufficiently low link density ($= L/\binom{N}{2}$) to avoid highly meshed topology. This is because the spreading patterns of our model will converge to that of contact-based model since in a fully meshed network, nodes can reach each other directly without intermediate nodes, resulting in a situation where the agents' traveling pathways coincide with contacts (*i.e.*, everyone is just one hop away from each other). Generation of SF graphs follows the preferential attachment process described in [65] where the probability of a new node attaches to a given node is proportional to the total node degree. The number of links to attach at each step, w is set to 2 as default.

For each graph, without loss of generality, we pre-compute the paths for each and every node pair using Dijkstra's algorithm, assuming non-weighted links. At time, $t = 0$, 10% of nodes in the network are randomly chosen to be the infected while the rest are assumed to be susceptible nodes. We then generate λN agents at each time step. Each agent is given a random origin and destination nodes and the agent moves along the pre-computed paths of the origin-destination node pair. The infection and node removal processes follows the mechanics described in Section 3.

We show in Figures 4 and 5 one representative result each for the instantaneous evolution of the infected, susceptible and removed fractions of population for $N = 100$ for random and scale-free graphs respectively.

Finally, we can also compute the prevalence of the epidemic (*i.e.*, the fraction of infected nodes) at time t in the network using the following:

$$\rho(t) = \frac{1}{N} \sum_{n=1}^N i_n(t). \quad (18)$$

We note that at steady-state, $\frac{di_n(t)}{dt}|_{t \rightarrow \infty} = 0$ and $i_{n\infty} = \lim_{t \rightarrow \infty} i_n(t) = 0$. Naturally, it follows that as $t \rightarrow \infty$, $\rho = 0$ and $\sum_n s_n + \sum_n r_n = N$.

Using the tuple $\{\beta, \mu, \lambda\} = \{0.5, 1.0, 2.0\}$ as the default setting, we show in Figure 6 for ER graphs and Fig. 7 for SF graphs that our model always predict the size of epidemic within the 99% confidence interval of the simulated results across topologies of different sizes. Further, we present in Fig. 8 and Fig. 9 the fraction of population infected over the lifetime of the epidemic. From these figures (also with 99% confidence interval), similar agreements between our model and simulation results can be observed when we vary the multiple of $p_c = \{2, 3, 4, 5, 6\}$ parameter for ER graphs and $w = \{2, 3, 4, 5, 6\}$ parameter for SF graphs. All confidence intervals were computed based on 100 simulation runs.

5 VULNERABILITIES FROM DIFFERENT PERSPECTIVES

In this section, we analyze the effect of epidemic spread by agents from three different perspectives. Specifically, we discuss (1) how vulnerable the network as a whole to our agent mobility-based epidemic, (2) vulnerability of individual nodes in the network and (3) agent's susceptibility to infection when traversing the network.

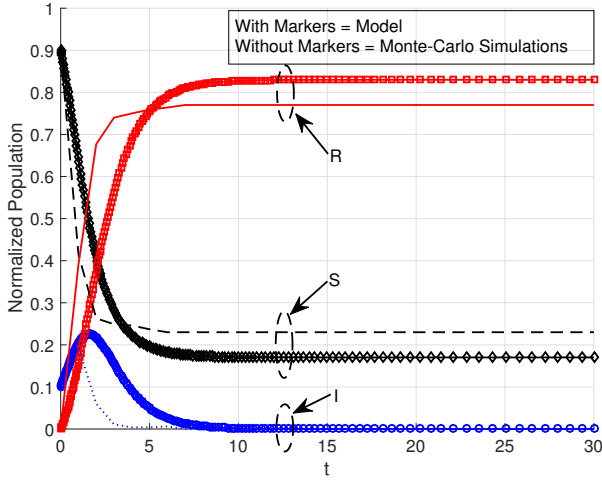


Fig. 4. (Color online) Instantaneous evolution of the population for a random graph of size $N = 100$ with $\lambda = 1$ for all nodes. Lines with markers are results from our model while solid, dashed and dotted lines without markers are results from Monte-Carlo simulations.

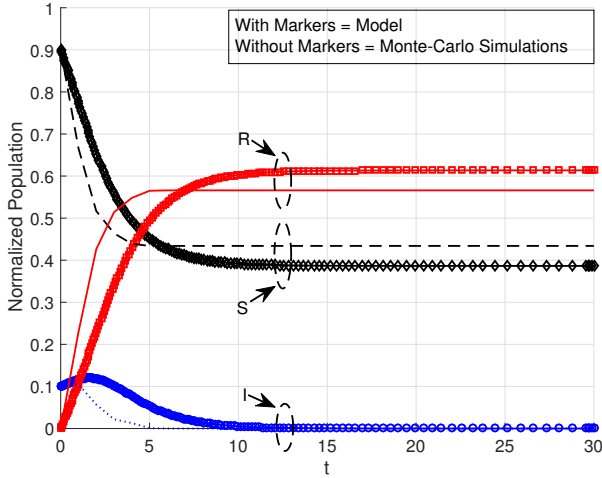


Fig. 5. (Color online) Instantaneous evolution of the population for a scale-free graph of size $N = 100$ with $\lambda = 1$ for all nodes. Lines with markers are results from our model while solid, dashed and dotted lines without markers are results from Monte-Carlo simulations.

5.1 Network as a whole

We derive some steady-state properties of our model for the network viewed as a whole system. We begin with the trivial observation that all agent mobility-based SIR epidemic will converge to an infection free state since an individual will only be infected for a finite time period after which it is no longer susceptible to re-infection. Without re-infection process, the number of individual susceptible to infection monotonically decreases while the number of individual gaining immunity monotonically increases and finally, results in the death of the epidemic, regardless of the amount and state of the agents within the network.

Often the most important question asked regarding an epidemic is whether there will be an outbreak (*i.e.*, will it spread or die off rapidly). A common quantity used in the literature to answer this question is the so-called epidemic

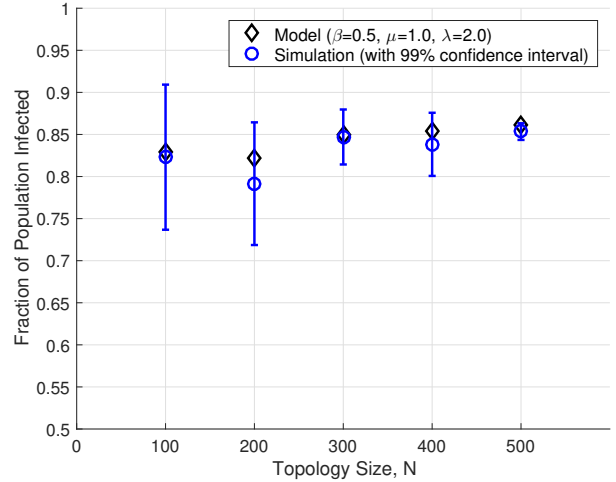


Fig. 6. (Color online) Epidemic size (*i.e.*, fraction of population who eventually contracted the infection) for a set of sample ER topologies with $N = \{100, 200, 300, 400, 500\}$ and $p = 2 \times p_c$.

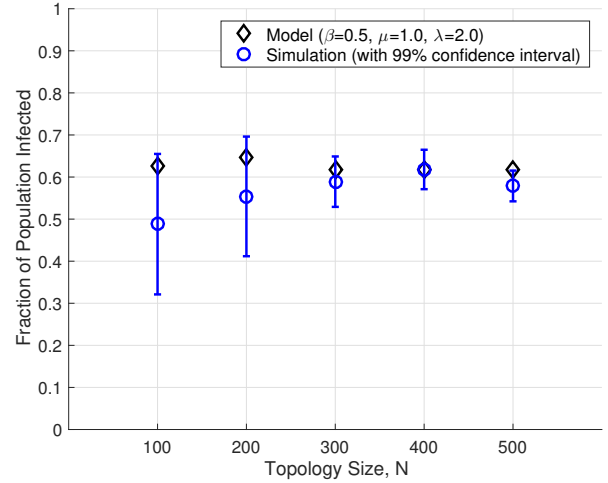


Fig. 7. (Color online) Epidemic size (*i.e.*, fraction of population who eventually contracted the infection) for a set of sample SF topologies with $N = \{100, 200, 300, 400, 500\}$ and $w = 2$.

threshold, τ_c , which defines the point of transition between an epidemic and a non-epidemic regime. If the effective spreading rate of the epidemic is below this threshold, then the epidemic will die off and vice versa. However, unlike the *SI* or *SIS* models⁷, the *SIR* model will always die off (*i.e.*, zero infection at the end since infected nodes will eventually become immune). We follow the definition given in [66] where the *SIR* epidemic threshold is defined as the point that separates between the network having a finite fraction of the population being infected and the network having infinitesimally small number of infected individuals in the limit of very large N .

⁷ In the *SI* model, there is no such transition point since all nodes will eventually be infected. In the *SIS* model, a node may be re-infected and thus, the network may converge to a steady-state such that the epidemic prevails over long period.

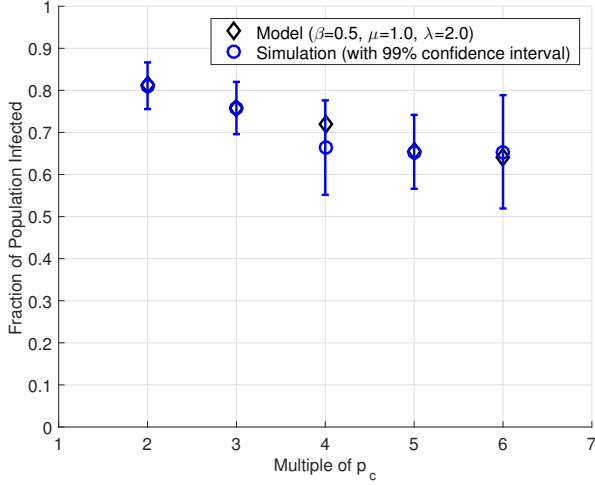


Fig. 8. (Color online) Epidemic size (*i.e.*, fraction of population who eventually contracted the infection) for a set of sample ER topologies with $N = 100$ and multiple of $p_c = \{2, 3, 4, 5, 6\}$.

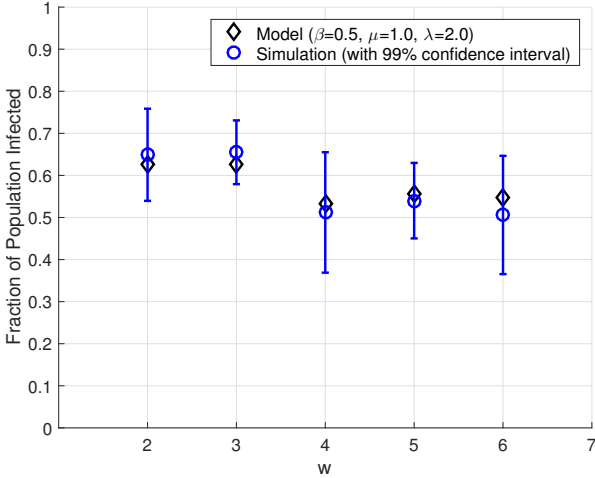


Fig. 9. (Color online) Epidemic size (*i.e.*, fraction of population who eventually contracted the infection) for a set of sample SF topologies with $N = 100$ and $w = \{2, 3, 4, 5, 6\}$.

Theorem 1 (Epidemic Threshold). *Given a network system characterized by matrix, C and the corresponding largest eigenvalue being γ_1 , then the epidemic threshold, τ_c is as follows:*

$$\tau_c = \frac{1}{\gamma_1} \quad (19)$$

where γ_1 is the largest eigenvalue of C .

Proof. Consider, at time $t = 0$, the initial infected fraction of population is very small (*i.e.*, $\lim_{N \rightarrow \infty} s_n(0) \approx 1; \forall n \in V$) and none of the individuals in the network is immuned ($r_n(0) = 0; \forall n \in V$). At such conditions, we can linearize Eq. 13 around the infection-free equilibrium as follows:

$$\frac{di_n(t)}{dt} = \beta \sum_{m=1}^N i_m(t) c_{n,m} - \mu i_n(t) \quad (20)$$

$$= \sum_{m=1}^N (\beta c_{n,m} - \mu \delta_{n,m}) i_m(t) \quad (21)$$

where the Kronecker delta, $\delta_{n,m} = 1$ when $n = m$.

Equivalently, we can re-write the equation in matrix form as follows:

$$\frac{dI(t)}{dt} = \beta H I(t) \quad (22)$$

where H is the $N \times N$ matrix

$$H = C - \frac{\mu}{\beta} \mathbb{1}. \quad (23)$$

where $\mathbb{1}$ is the identity matrix.

By using the principle of linear stability for systems of ODEs, we can observe that matrix H (see Eq. 23), by construction, has the same spectrum as C but shifted to the left by μ/β . Thus, the largest eigenvalue of H , σ_1 can be written as follows:

$$\sigma_1 = \gamma_1 - \frac{\mu}{\beta} \quad (24)$$

where γ_1 is the largest eigenvalue of C as we adopt the convention $\gamma_1 > |\gamma_2| \geq |\gamma_3| \geq \dots \geq |\gamma_N|$ where $|\cdot|$ denotes the absolute value if the eigenvalue is real or modulus if it is complex.

The epidemic threshold is given by $\sigma_1 = 0$ and from Eq. 24, we can then state that the epidemic threshold as:

$$\tau_c = \frac{\beta}{\mu} = \frac{1}{\gamma_1}. \quad (25)$$

Therefore, whether there is an outbreak of epidemic is determined by γ_1 whereby the epidemic is easier to spread when γ_1 is large. This is logical since we get higher γ_1 when $c_{n,m}$ assumes larger values (see Theorem 2 in [25]) which means there is higher volume of agents traversing nodes (and thus, passing on the contagion) within the network. \square

While C is almost surely non-symmetric, it remains by definition to be a non-negative square matrix. Using this property, we can then state the following.

Theorem 2 (Upper and lower bounds of τ_c). *Given an agent mobility induced epidemic characterized by matrix C , its epidemic threshold has the following upper and lower bounds:*

$$\frac{1}{\max_{1 \leq n \leq N} b_n(C)} \leq \tau_c \leq \frac{1}{\min_{1 \leq n \leq N} b_n(C)} \quad (26)$$

where $b_n(C)$ is the n -th average 2-row sum of C and defined as below:

$$b_n(C) = \frac{\sum_{k=1}^N (c_{n,k} \sum_{m=1}^N c_{k,m})}{\sum_{m=1}^N c_{n,m}} \quad (27)$$

Proof. We use the result from [68] which refined Frobenius's well-known lemma on largest eigenvalue and stated that the largest eigenvalue of an $N \times N$ non-negative matrix to be bounded to the following:

$$\min_{1 \leq n \leq N} b_n(C) \leq \gamma_1 \leq \max_{1 \leq n \leq N} b_n(C). \quad (28)$$

Using Eq. 28 above in combination of Eq. 19, we then get Eq. 26. \square

From Theorem 2, we see that τ_c is upper and lower bounded by the nodes traversed by the highest and lowest weighted average number of agents in the network respectively. As such, we can also say that the more uniformly distributed the agent traffic is amongst the network nodes, the tighter the bounds are, which makes intuitive physical sense.

A special case is when all nodes are traversed by equal number of agents, resulting in a symmetric hollow matrix C (i.e., $c_{n,n} = 0; \forall n \in N$) with all off diagonal elements (i.e., $c_{n,m}; n \neq m$) equal in value. Under such conditions, we observe C 's largest eigenvalue, $\gamma_1 = (N - 1)c_{n,m}$ while the rest of the eigenvalues are $-c_{n,m}$ (i.e., with multiplicity $N - 1$). In this case, the epidemic threshold is simply as follows:

$$\tau_c = \frac{1}{(N - 1)c_{n,m}}; \forall n, m \in N : n \neq m \quad (29)$$

except for $N = 1$, when $\tau_c = 1/c_{1,1}$ which is ∞ due to the assumption that there is no self traffic (i.e., in this case, no agents). Trivially, the sum of all eigenvalues for this case is zero (i.e., $\sum_n \gamma_n = \sum_n c_{n,n} = 0$).

5.2 Individual Node Vulnerability

In the previous section, we obtained the epidemic threshold being inversely proportional to the largest eigenvalue of C via eigen decomposition of matrix C (see Theorem 1). Another useful product of this decomposition relates to the principal eigenvector of matrix C .

Definition 1 (Node vulnerability). *The vulnerability of node n is proportional to $(z_1)_n$ where $(z_1)_n$ be the n -th element of the principal eigenvector corresponding to the largest eigenvalue of matrix C , γ_1 .*

As prior mentioned, we consider undirected connected $G(V, E)$. Thus, there is a path between all node pairs. Assuming positive interactions between nodes (i.e., $\lambda_{k,n} > 0; \forall k \neq n$), matrix C is irreducible. As such, in our case, the Perron-Frobenius theorem asserts that the eigenvector corresponding to the largest eigenvalue can be chosen to have strictly positive components.

By eigen decomposition of matrix C , we get the following:

$$\gamma_1(z_1)_n = \sum_{m=1}^N c_{n,m}(z_1)_m \quad (30)$$

By observing Eq. 30, we see that $(z_1)_n$ is proportional to the row sum of C which in turn reflects the total number of agents traversing or destined to node n . Analogous to the concept of eigenvector centrality [69], we can see that in its canonical form, matrix C also allow us to identify the nodes that are most vulnerable to the epidemic spread, i.e., nodes with highest $(z_1)_n$ being the ones most prone to being infected.

This gives us a very useful tool. Using the principal eigenvector of C , we can now "harmonize" the vulnerability of all the nodes in the network (i.e., all nodes equally

susceptible to being infected) by exploiting the fact that γ' must be an eigenvalue of C if the row sum of C equals γ' . To achieve this, we simply need to distribute the agent traffic across all nodes uniformly. The solution will be dependent on A as well as how much control one has over the traveling agents.

5.3 Agents' Vulnerability

In our model, we have bi-directional infection. An agent may infect a node and vice versa, an infectious node can also infect an agent. In some scenarios (e.g., transportation), the agents are actual human beings. As such, it is of interest to understand how prone the agents are in the system to infection when traversing within the network. Specifically, we are interested in the following question: What is the likelihood for an agent originated from a node, v_m destined to node, v_n (in this case, a node is a site; e.g., airport, train station, etc.) catching an infection within its journey?

Let $X_{n,m}(t)$ be the final state of an agent of interest, $g_{n,m}$, traveling from v_m to v_n . For our model, $X_{n,m}(t) = \{S, I\}$, i.e., an agent can only be either susceptible or infected at any one time.

Solving the system of equations of our model (i.e., Eqs. 12, 13 and 14), we can obtain the infected probability of each node individually (i.e., $i_n(t)$). Using this result from our model, we can now compute the probability of $g_{n,m}$ be in the infected state as

$$Pr[X_{n,m}(t) = I] = 1 - \prod_{k=1}^N (1 - i_k(t)\mathbf{1}_{k \in P_{n,m}}) \quad (31)$$

In other words, given the traveling route of any agent (i.e., $P_{n,m}$), we can use Eq. 31 to predict the probability of the agent getting infected by the end of its journey.

Furthermore, based on Eq. 31, the expected number of agents travelling from v_n to v_m that gets infected is simply $\lambda_{n,m}Pr[X_{n,m}(t) = I]$. Using this, the total agents getting infection in the network can be computed as follows:

$$\sum_{n=1; n \neq m}^N \sum_{m=1; m \neq n}^N \lambda_{n,m}Pr[X_{n,m}(t) = I] \quad (32)$$

Equation 32 can be considered as the first-order estimation since in our method, we have assumed no agent traveling delays (see Section 3) to reduce the analysis complexity. Summing Eq. 32 over the time since outbreak until the death of the epidemic will give us the estimation of total agent population being infected in the entire epidemic. Usually, the number of agents in a system is very large compared to the number of nodes of the network. As such, it is usually intractable to model the state transition of each agent individually. For instance, in our case study, the agents recorded is approximately 3 billions which is about 10^7 order of magnitude greater than the number of nodes (see Section 6.1).

6 CASE STUDY

In the following, we compare our agent mobility-based SIR epidemic model with conventional contact-based epidemic via a hypothetical infection spreading in the London Underground network with the commuters being the agents carrying the infection.

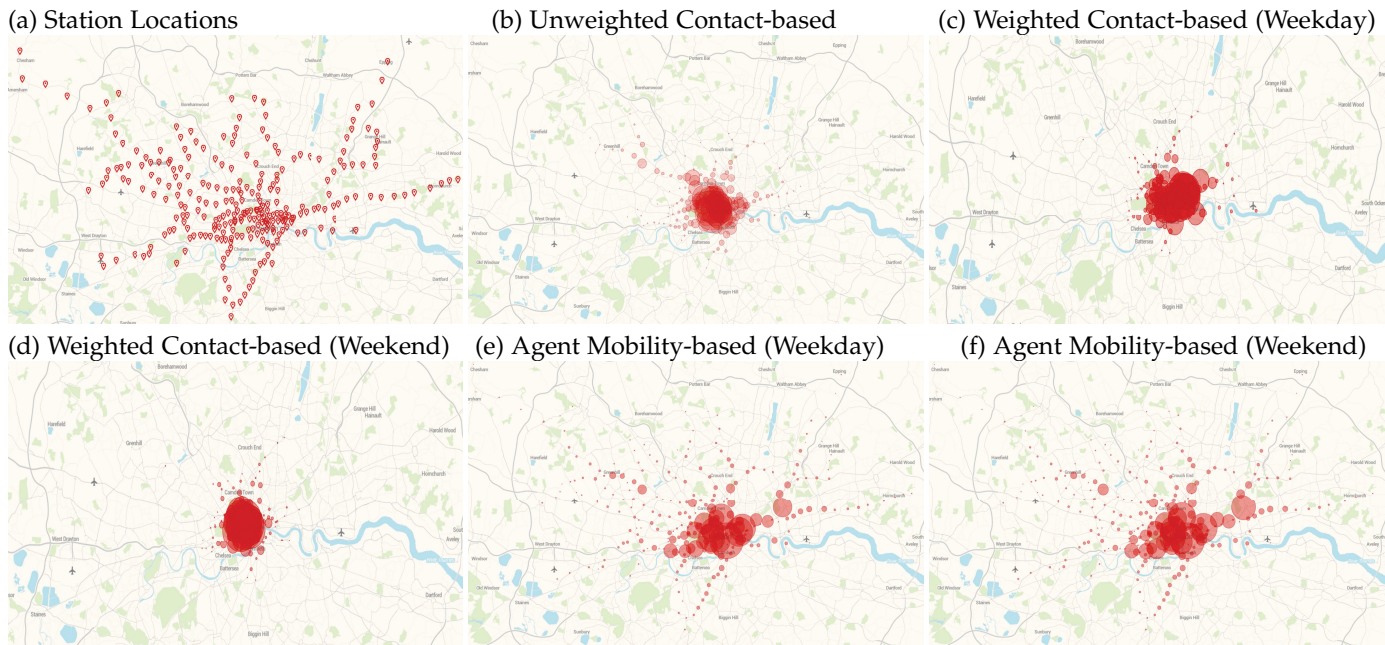


Fig. 10. (Color online) Comparison of contact-based and agent mobility-based epidemic spreading in the London Underground network system: (a) Geographical locations of London Underground tube stations, (b) Epidemic spreading based on conventional contact-based SIR model, (c) Weighted contact-based SIR model based on weekday data, (d) Weighted contact-based SIR model based on weekend data, (e) Epidemic spreading caused by agents travelling in the network system based on weekday data and (f) Epidemic spreading caused by agents travelling in the network system based on weekend data. The size of the red circles are proportional to how prone the stations are to being infected.

6.1 Dataset

For this case study, we use a dataset provided by Transport for London (TfL)⁸ which recorded annual number of commuters entering and exiting individual London Underground stations⁹ from 2007 to 2014. The dataset also separately record the entries for weekdays and weekends. We complement the dataset with the London Underground map. We show the geographical locations of all considered Underground stations in Fig. 10(a) based on their longitude and latitude data. Hereafter, we represent the London network by the adjacency matrix, A_{tube} . For our purpose, we use the latest data of the year 2014 with a total of 2.8956 billion observations. We note that qualitatively, the yearly activity is rather similar per station. We restrict ourselves to consider only Underground stations – *i.e.*, tram, overground and Dockland Light Railway (DLR) and train stations are excluded. We have thus a total of $N_{tube} = 265$ stations. Figure 11 shows the degree distribution of the London Underground network in which we see many stations acting as bridge nodes (*i.e.*, with degree = 2) sharing similar features as radial power grid networks (see for example [9]). The London Underground network is also divided into 9 zones where the station density is highest in Zone 1 (*i.e.*, in the city center) as well as 11 different tube lines.

In addition, we note that the exact paths the commuters took after entering the Underground network are not recorded. As such, we assume that the commuters are rational (*i.e.*, they take the shortest available path to reach their destinations). We also further assume that the commuters

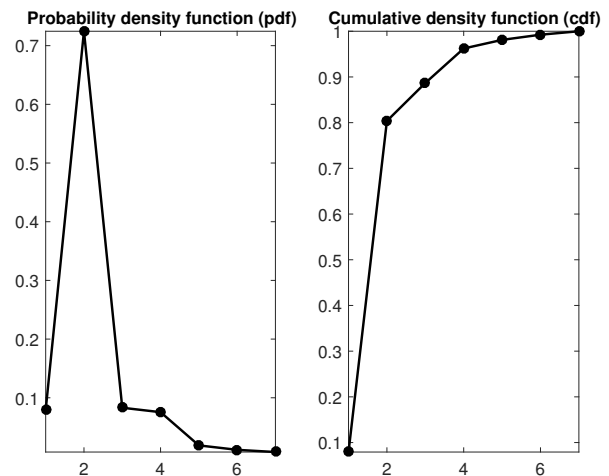


Fig. 11. Degree distribution of London Underground tube stations in the dataset showing majority of stations have low connectivity.

do not make errors when traveling which is obviously a simplifying assumption since this excludes those often occur incidents when commuters (especially tourists) that changed tube at the wrong platform or mistakenly took the tube in the opposite direction, *etc.*

6.2 Analysis and Observations

Using the above complemented dataset, we compute and show in Fig. 10 both the epidemic spreading based on conventional contact-based SIR model and based on our model in which infection is spread via carrier agents traveling in the network system. Fig. 10(a) shows the actual

8. <https://tfl.gov.uk/>

9. London Underground is also commonly known as the “tube”. We use both terms interchangeably.

geographical locations of the London Underground tube stations which are distributed across the figure. Fig. 10(b) shows contact-based SIR model computed using the method presented in [12] (indicated with “Unweighted”). Furthermore, we also consider a version of weighted contact-based SIR model where each link from a station is weighted by the commuter volume based on the dataset. These are illustrated in Fig. 10(c) and (d) for commuter traffic data on weekday and weekend respectively. Finally and correspondingly, we compute and show the spread using our agent mobility-based model for both weekday and weekend data in Fig. 10(e) and (f). We use red circles to indicate the vulnerability of the Underground stations: the greater the radius, the more vulnerable the node is. The first observation when comparing Fig. 10(b)–(d) and 10(e)–(f) is that our model predicts a *wider* spatial spread of the epidemic towards the outskirts of the city while the contact-based epidemic shows infection highly concentrated in the city center with peripheral stations almost immune to the infection.

We then focus our attention to the city center (*i.e.*, mostly Underground stations within Zone 1). From Fig. 12, it is clear that, for contact-based epidemic, underground stations located within the city center are predicted to be highly susceptible to infection. Unweighted contact-based epidemic model only takes into account the topology of the network and hence, stations are deemed to be more vulnerable by virtue of being in the *center* of the network. Meanwhile, by additionally taking into account the activity level each tube station, weighted contact-based SIR model further reflects in its prediction the impact of station’s activity on the vulnerability of the station’s immediate neighbours but not the entire route of the commuters’ journeys. In contrast, for our agent mobility-based epidemic model, we see many *central* stations in Zone 1 are less susceptible to being contracted with infection (see Fig. 13). From the figures (highlighted by the dashed circle and oval in Fig. 12 and Fig. 13 respectively), we also see that the contact-based spreading is omnidirectional while in our case, the spreading follows the tube lines (*i.e.*, agents’ travel routes).

Next, we illustrate in Fig. 14 that unweighted contact-based epidemic model shows linear correlation between distance of station location from city center and their vulnerability, *i.e.*, stations located nearer to central London are proportionally more vulnerable to infection. Such is not the case for our model where the susceptibility of stations in a tube line is not only dependent on their geographical locations but also the number of agents travelling across them. As such, we observe from the figure many tube stations in Zone 1 are actually not highly prone to infection while some stations outside of Zone 1 have high susceptibility (see Fig. 10(e)–(f)).

We further show the predicted vulnerability of two main tube lines, *i.e.*, the Northern line (see Fig. 15) which connects north and south London (*i.e.*, vertically across the London city) and the Central line (see Fig. 16) which runs across west and east London (*i.e.*, horizontally across the London city). Both tube lines go through the city center. From the figures, we see that contact-based model’s predictions exhibit clear bias against stations located in central London (stations in Zone 1 of the network are almost always more prone to

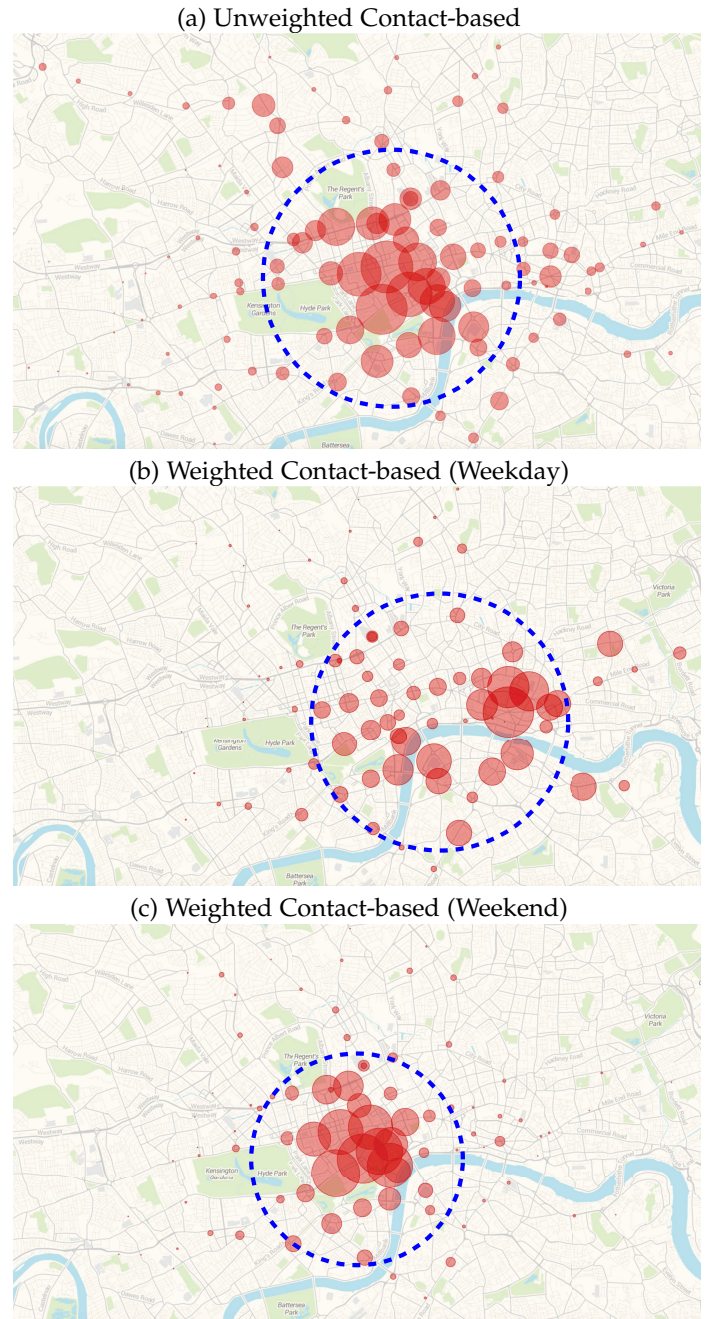


Fig. 12. (Color Online) Weighted and unweighted contact-based epidemic spreading within and around Zone 1 of the London Tube network.

infection). By considering the commuters’ mobility within the network, our model shows otherwise – *e.g.*, from Fig. 15, stations 14-18 in the Northern line are not highly susceptible to infection despite being located in Zone 1. Our model shows that station susceptibility is not solely based on locations, resulting in multiple spikes – *e.g.*, Fig. 16 shows some Central line stations in Zone 3 are also being highly vulnerable.

Finally, we highlight the inflexibility and insensitivity of contact-based epidemic model with Table 2. First of all, while all the most vulnerable stations are located within Zone 1 of the London Underground network, we see the Stratford station which is located in Zone 3 being ranked as

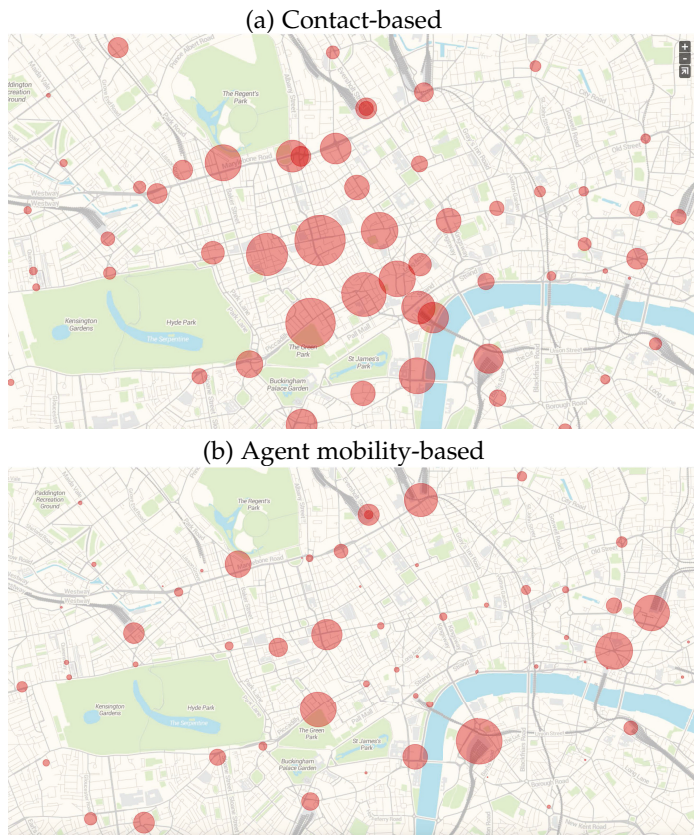
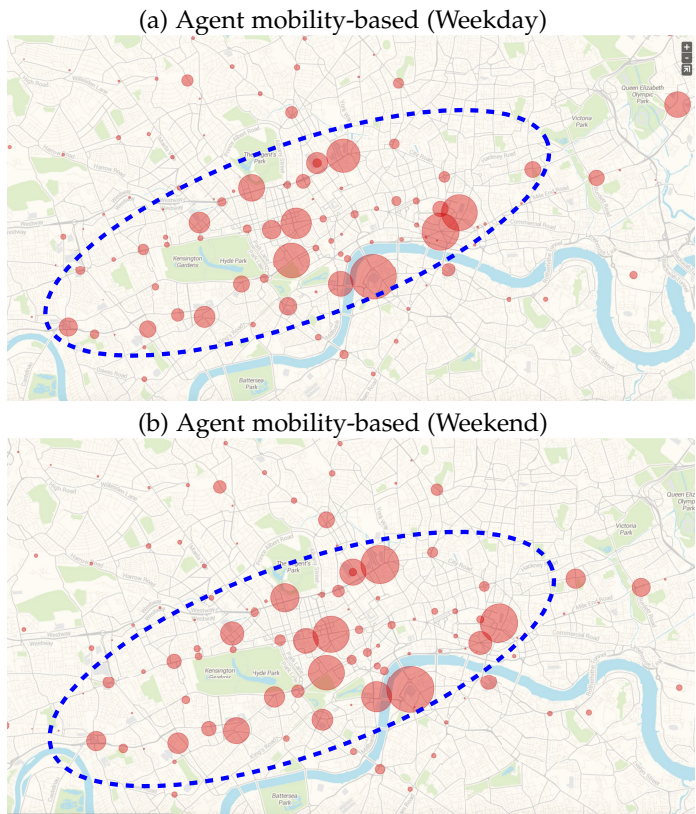


Fig. 13. (Color Online) Agent mobility-based model is sensitive to the infection carriers' travelling paths and stations may not be prone to infection even when geographically located at the center of the city.

Fig. 14. (Color Online) For unweighted contact-based epidemic model, the vulnerability of stations along a tube line gradually increases when moving towards the city center. On the other hand, our model shows that the vulnerability of stations along a tube line is not linear and dependent on the level of activities at each station. A tube station at the city center may not necessarily be the most vulnerable site for infection.

one of the most susceptible tube station by our model. This station is directly linked to Stratford International train station that services the southeastern high-speed line. Second, our model identified that London Waterloo being one of the busiest tube stations in the network is highly susceptible to infection. Third, due to the fact that the Waterloo and City line is closed on Sundays, the activities at the Bank station is significantly reduced over the weekend. This is reflected by our model. The above are not captured in the contact-based model.

Northern line underground stations from south to north London

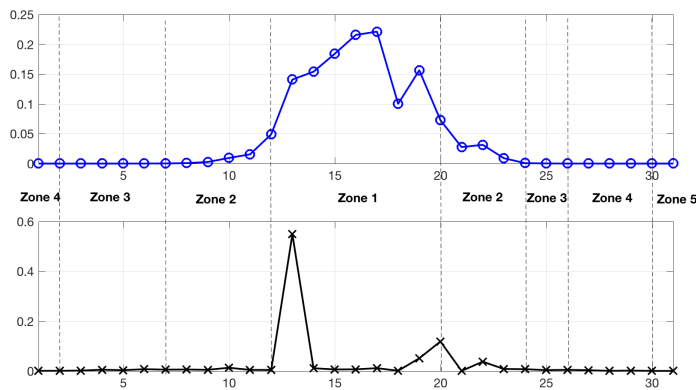


Fig. 15. (Color Online) Comparing station vulnerability along the Northern line tube stations from south to north London: (top) Unweighted contact-based model and (bottom) agent mobility-based model.

TABLE 2

Top 10 most vulnerable London Underground stations ranked in descending order based on contact-based model (Column 1) and agent mobility-based model (Column 2 and 3).¹⁰

Contact-based	Mobility-based (Weekday)	Mobility-based (Weekend)
Oxford Circus (6)	Waterloo (5)	Waterloo (5)
Green Park (6)	Bank (5)	Liverpool Street (5)
Piccadilly Circus (4)	Green Park (6)	King's Cross (7)
Bond Street (4)	Liverpool Street (5)	Oxford Circus (6)
Tottenham Court Road (4)	King's Cross (7)	Green Park (6)
Leicester Square (4)	Oxford Circus (6)	Stratford (3)
Baker Street (7)	Baker Street (7)	Westminster (4)
Westminster (4)	Stratford (3)	Baker Street (7)
Charing Cross (3)	Westminster (4)	Euston (4)
Regent's Park (2)	Euston (4)	South Kensington (3)

10. Bracketed values indicate the node degrees.

7 CONCLUSIONS

Various spreading phenomena in real world spanning across natural ensembles and human engineered networks rely on agents within the system to act as the physical transport medium of infection. Such phenomena is the subject of this paper where we model epidemic spreading caused by infectious agents. Our modelling methodology is general in nature and is applicable to different context and

Central line underground stations from west to east
London

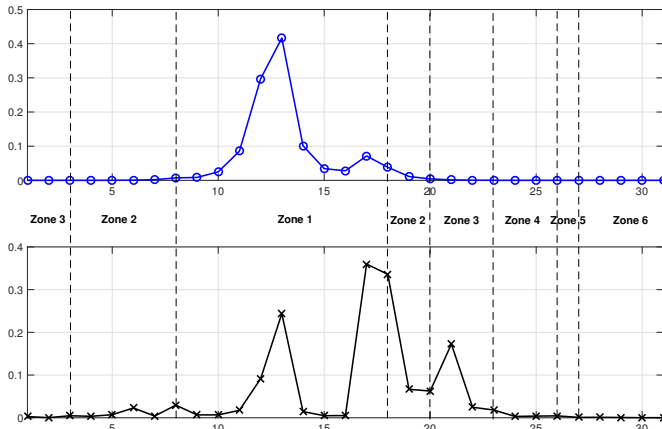


Fig. 16. (Color Online) Comparing station vulnerability along the Central line tube stations from south to north London: (top) Unweighted contact-based model and (bottom) agent mobility-based model.

network types since our interest here is not to solve specific problem but to illustrate the analysis of diffusion processes caused by mobile agents within the system. Our analytical framework consider explicitly individual node within the network rather than the network as a whole. An artefact of this approach is that our solution reduces the complexity from exponential of the exact solution to polynomial complexity. In our approach, we define a characterization matrix, C , which contains the information regarding the number and the mobility pattern of the agents in the system. We derive the epidemic threshold to be the reciprocal of matrix C 's largest eigenvalue. Furthermore, by considering each node individually, our approach provides us infection susceptibility information on the nodal-level and allows direct inference of level of vulnerability of each node in the network. This facilitates immunization or protection schemes to be deployed in the system (depending on the controllability of the agents as well as the possibility to alter network structure (via topology change effecting adjacency matrix A and consequently, matrix C)). For instance, this can be an islanding operation in a power grid which sections off parts of a grid and redirecting the power flow from alternative primary sub-station(s) to avoid cascading failures. Furthermore, we also take the viewpoint from the agents' perspective and derived the first order approximation of agents' vulnerability. We first show agreements of predictions based on our model against Monte Carlo simulations. Then, we investigated and compared the spreading process effected by contacts and by agent mobility with a realistic case study using dataset recording the activity of London Underground commuters, one of the busiest metro system in the world. We found that due to the mobility of agents, the spreading radius of an epidemic is larger than that predicted for contact-based epidemic models. We also found that with our model, node susceptibility no longer directly depends on its location in the network. While in contact-based epidemics, node proneness to infection is directly proportional to its location relative to the *center* of the network, our model does not exhibit such correlation. Instead, the vulnerability of the node is dependent on the agents'

mobility pattern (*i.e.*, edge nodes may be highly susceptible to infection if high number of agents traverse across or destine to them). These dynamics due to agents' activities are not captured by conventional contact-based epidemic models.

ACKNOWLEDGMENT

This work was partially funded by the Engineering and Physical Sciences Research Council (EPSRC) under the CHIST-ERA CONCERT (A Context-Adaptive Content Ecosystem Under Uncertainty), project number *I1402*.

REFERENCES

- [1] R. M. Anderson and R. M. May, "Infectious Diseases of Humans", Oxford: Oxford University Press, 1991.
- [2] F. Darabi Sahneh and C. Scoglio, "Epidemic spread in human networks," IEEE Conference on Decision and Control, Dec. 2011.
- [3] Y. Iturria-Medina, *et al.*, "Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders," PLoS Computational Biology, 10(11), e1003956, 2014.
- [4] J. O. Kephart and S. R. White, "Measuring and modelling computer virus prevalence," IEEE Symposium on Security and Privacy, 1993.
- [5] M. Garetto, W. Gong and D. Towsley, "Modeling malware spreading dynamics," IEEE INFOCOM, vol. 3, pp. 1869-1879, 2003.
- [6] T. N. Dinh, *et al.*, "Cost-effective viral marketing for time-critical campaigns in large-scale social networks," IEEE/ACM Transactions on Networking, vol. 22, no. 6, pp. 2001-2011, December 2014.
- [7] S. V. Buldyrev, *et al.*, "Catastrophic cascade of failures in interdependent networks," Nature 464(7291), pp. 1025-1028, 15 April 2010.
- [8] A. Bernstein, *et al.*, "Power Grid Vulnerability to Geographically Correlated Failures - Analysis and Control Implications," IEEE INFOCOM, pp. 2634-2642, 2014.
- [9] W. K. Chai, *et al.*, "Resilience of Interdependent Communication and Power Distribution Networks against Cascading Failures," 15th IFIP Networking, Vienna, Austria, May 2016
- [10] D. J. Daley and J. Gani, "Epidemic modelling: an introduction," Cambridge, UK, Cambridge Press, 1999.
- [11] Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos, "Epidemic spreading in real networks: an eigenvalue viewpoint," 22nd Int'l. Symp. Reliable Distributed Systems (SRDS), Oct. 2003, pp. 25-34.
- [12] P. Van Mieghem, J. Omic and R. Kooij, "Virus spread in networks," IEEE/ACM Transactions on Networking, vol. 17, no. 1, Feb. 2009.
- [13] P. Van Mieghem, "The N -intertwined SIS epidemic network model," Computing, 93(2):147-169, 2011.
- [14] D. Centola and M. Macy. "Complex Contagions and the Weakness of Long Ties." American Journal of Sociology, vol. 113, no. 3, 2007.
- [15] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics", Proc. Royal Society A: Mathematical, Physical and Engineering Sciences 115 (772): 700.
- [16] Z. Ruan, P. Hui, H. Lin and Z. Liu, "Risks of an epidemic in a two-layered railway-local area traveling network", the European Physical Journal B, 86:13, 2013.
- [17] V. Colizza, A. Barrat, M. Barthélemy and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics", Proceedings of National Academy Science, USA, pp. 2015-2020, 2006.
- [18] M. C. Gonzalez, C. A. Hidalgo and A. Barabasi. "Understanding individual human mobility patterns," Nature 453(7196), 2008.
- [19] C. Song, Z. Qu, N. Blumm, and A. Barabasi, "Limits of predictability in human mobility", Science, 2010.
- [20] L. Carettoni, C. Merloni and S. Zanero, "Studying bluetooth malware propagation: The bluebag project," IEEE Security and Privacy 5, 2 (2007), 17-25.
- [21] P. Wang, M. Gonzalez, C. Hidalgo and A. Barabasi, "Understanding the spreading patterns of mobile phone viruses," Science 324, 5930 (2009), 1071.
- [22] G. Xylomenos, *et al.*, "A Survey of Information-Centric Networking Research," in Communications Surveys & Tutorials, IEEE , vol. 16, no. 2, pp.1024-1049, 2nd Quarter 2014.
- [23] W. K. Chai, D. He, I. Psaras and G. Pavlou, "Cache "less for more" in information-centric networks (Extended version)," Computer Communications, vol. 36, no. 7, pp. 758-770, 2013.

- [24] V. Jacobson, *et al.*, "Networking Named Content," 5th ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT); Rome, Italy, 2009.
- [25] W. K. Chai and G. Pavlou, "Path-based epidemic spreading in networks," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, Feb. 2017.
- [26] K. Drakopoulos, A. Ozdaglar and J. Tsitsiklis, "An efficient curing policy for epidemics on graphs," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 2, pp. 67-75, Jul-Dec 2014.
- [27] K. Scaman, A. Kalogeratos and N. Vayatis, "Suppressing epidemics in networks using priority planning," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 4, pp. 271-285, 2016.
- [28] Y. Li, *et al.*, "Optimal distributed malware defense in mobile networks with heterogeneous devices," *IEEE Transactions on Mobile Computing*, vol. 13, no. 2, pp. 377-391, Feb. 2014.
- [29] S. Eshghi, *et al.*, "Optimal patching in clustered malware epidemics," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 283-298, Feb. 2016.
- [30] Z. Zhang, *et al.*, "Modeling epidemics spreading on social contact networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp. 410-419, Sept. 2015.
- [31] I. Gupta, A.-M. Kerमारrec and A. J. Ganesh, "Efficient and adaptive epidemic-style protocols for reliable and scalable multicast," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 7, pp. 593-605, July 2006.
- [32] D. Chakrabarti, *et al.*, "Information survival threshold in sensor and p2p networks," *IEEE INFOCOM*, pp. 1316-1324, 2007.
- [33] M. Vojnović, V. Gupta, T. Karagiannis and C. Gkantsidis, "Sampling strategies for epidemic-style information dissemination," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, August 2010.
- [34] Y. Li and J. C. S. Lui, "Epidemic attacks in network-coding-enabled wireless mesh networks: detection, identification, and evaluation," *IEEE/ACM Transactions on Networking*, vol. 12, no. 11, pp. 2219-2231, November 2013.
- [35] M. Ogura and V. Preciado, "Stability of Spreading Processes over Time-Varying Large-Scale Networks," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 1, pp. 44-57, 2016.
- [36] P. Basaras, D. Katsaros, and L. Tassiulas, "Detecting influential spreaders in complex, dynamic networks," *IEEE Computer*, vol. 46, no. 4, pp. 24-29, Apr. 2013.
- [37] D. Bernoulli, "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir," *Mém. Math. Phys. Acad. Roy. Sci.*, Paris, 1-45, 1760.
- [38] A. L. Barabasi, "Network Science," Cambridge University Press, United Kingdom, 2016.
- [39] R. M. May and A. L. Lloyd, "Infection dynamics on scale-free networks," *Physical Review E*, vol. 64, 066112, 2001.
- [40] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys Rev Lett*, vol. 86, no. 14, 2001.
- [41] M. Boguñá, R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in complex networks with degree correlations," 18th Conference on Statistical Mechanics, 2003.
- [42] Y. Moreno, R. Pastor-Satorras and A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks," *Eur. Phys. J. B.* no. 26, pp. 521-529, 2002.
- [43] M. E. J. Newman, "The spread of epidemic disease on networks," *Physics Review E* 66, 016128, 2002.
- [44] M. J. Keeling and P. Rohani, "Modeling infectious diseases in humans and animals," Princeton University Press, 2008.
- [45] M. Keeling, D. Rand and A. Morris, "Correlation models for childhood epidemics," *Proc. Roy. Soc. Lond. B* 264 1149-1156, 1997.
- [46] M. J. Keeling, "The effects of local spatial structure on epidemiological invasions," *Proc. Roy. Soc. Lond. B* 266 859-867 1999.
- [47] K. T. D. Eames and M. J. Keeling, "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases," *Proc. Natl. Acad. Sci. U.S.A.* 99 13330-13335 2002.
- [48] Z. Nikoloski, N. Deo and L. Kucera, "Correlation model of worm propagation on scale-free networks," *Complex Network Modelling*, vol. 3, pp. 169-182, 2006.
- [49] M. Youssef and C. Scoglio, "An individual-based approach to SIR epidemics in contact networks," *Journal of Theoretical Biology*, vol. 283, no. 1, pp. 136-144, 2011.
- [50] V. Colizza and A. Vespignani, "Invasion threshold in heterogeneous metapopulation networks," *Physical Review Letters*, 99, 148701, 2007.
- [51] V. Belik, T. Geisel and D. Brockmann, "Natural human mobility patterns and spatial spread of infectious diseases," *Physical Review X*, 1, 011001 (2011).
- [52] D. Balcan and A. Vespignani, "Phase transitions in contagion processes mediated by recurrent mobility patterns," *Nature Physics* 7.7, 581-586 (2011).
- [53] V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," *Journal of Theoretical Biology* 251, pp. 450-467, 2008.
- [54] D. Balcan, *et al.*, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proc. National Academy of Sciences (PNAS)*, vol. 106, no. 51, 21484-21489, 2009.
- [55] P. Bajardi, *et al.*, "Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic," *PLoS One* vol. 6, no. 1, e16591, 2011.
- [56] S. Meloni, A. Arenas and Y. Moreno, "Traffic-driven epidemic spreading in finite-size scale-free networks," *Proc. National Academy of Sciences (PNAS)*, vol. 106, no. 40, 16897-16902.
- [57] H.-X. Yang, Z.-X. Wu and B.-H. Wang, "Suppressing traffic-driven epidemic spreading by edge-removal strategies," *Physical Review E* 87, 064801, 2013.
- [58] H.-X. Yang and Z.-X. Wu, "Suppressing traffic-driven epidemic spreading by use of the efficient routing protocol," *Journal of Statistical Mechanics: Theory and Experiment*, no. 3, 1742-5468/14/P03018.
- [59] M. Ajelli, *et al.* "Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models," *BMC Infectious Diseases*, 10:190, 2010.
- [60] F. Darabi Sahneh, C. Scoglio and P. Van Mieghem, "Generalized epidemic mean-field model for spreading processes over multi-layer complex networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1609-1620, Oct. 2013.
- [61] P. Tume and M. Roughan, "Internet traffic matrices: A primer," *Recent Advances in Networking*, Vol. 1. ACM SIGCOMM, pp. 108-163, August 2013.
- [62] Cisco White Paper, "Building accurate traffic matrices with demand deduction," http://www.cisco.com/c/en/us/products/collateral/routers/wan-automation-engine/white_paper_c11-728552.pdf, 2013. Last accessed: 11-November-2015.
- [63] P. Van Mieghem, "Performance analysis of Complex networks and systems," Cambridge, Cambridge University Press, 2014.
- [64] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae* no. 6, pp. 290-297, 1959.
- [65] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509-512, Oct. 1999.
- [66] M. Boguñá, R. Pastor-Satorras and A. Vespignani, "Absence of epidemic threshold in scale-free networks with degree correlations," *Physical Review Letters*, vol. 90, no. 2, 17 January 2003.
- [67] A. Berman, R. J. Plemmons, "Nonnegative matrices in the mathematical sciences," Society of Industrial Mathematics, 1994.
- [68] X. Zhang and J. Li, "Spectral radius of non-negative matrices and digraphs," *Acta Mathematica Sinica*, vol. 18, no. 2, pp. 293-300, April 2002.
- [69] S. Wassermann and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge, Cambridge University Press, 1994.



Wei Koong Chai received the B.Eng. degree in electrical engineering from the Universiti Teknologi Malaysia, Johor Bahru, Malaysia, in 2000, and both the M.Sc. (Distinction) and the Ph.D. degrees from the University of Surrey, Surrey, U.K., in 2002 and 2008, respectively. He is currently a Senior Lecturer in the Department of Computing and Informatics, Bournemouth University, Dorset, U.K. as well as a Visiting Academic – Honorary Senior Research Associate – at University College London (UCL). Prior to this,

he was with the Department of Electronic and Electrical Engineering, UCL, first since 2008 as Research Fellow and subsequently since 2013 as Senior Research Fellow. His research spans across heterogeneous networks including wired/wireless networks and cyber physical systems. His current research interests include information-centric networking, smart grid communication and network science. He has also in the past involved in research on quality of service, resource management (e.g., for satellite networks and wireless mesh networks), cross-layer design (specifically on interaction of protocols at different layers), traffic engineering, and network optimization.