Inferring Contagion Patterns in Social Contact Networks Using a Maximum Likelihood Approach

Lauren M. Gardner[1]; David Fajardo[2]; and S. Travis Waller[3]

Abstract: The spread of infectious disease is an inherently stochastic process. As such, real-time control and prediction methods present a significant challenge. For diseases that spread through direct human interaction, the contagion process can be modeled on a social contact network where individuals are represented as nodes, and contact between individuals is represented as links. The objective of the model described in this paper is to infer the most likely path of infection through a contact network for an ongoing outbreak. The problem is formulated as a linear integer program. Specific properties of the problem are exploited to develop a much more efficient solution method than solving the linear program directly. The model output can provide insight into future epidemic outbreak patterns and aid in the development of intervention strategies. The model is evaluated for a combination of network structures and sizes, as well as various disease properties and potential human error in assessing these properties. The model performance varies based on these parameters, but it is shown to perform best for heterogeneous networks, which are consistent with many real world systems. **DOI: 10.1061/(ASCE)NH.1527-6996.0000135.**

Author keywords: Network optimization; Contagion models; Social contact networks.

Introduction

Many factors contribute to the spread and control of disease within a region, such as the standard of living, infection prevention practices (e.g., vaccination), local programs (e.g., health, emergency response), and, also critically significant, the interaction patterns among individuals. Today, a large proportion of the population lives in increasingly dense conditions, an ideal environment for rapid disease transmission. The stochastic nature of the contagion process (i.e., contact between an infectious and susceptible person may or may not result in a new infection) makes it difficult to identify the path of infection or predict the impact that a new disease might have on a region.

Over the last 100 years, significant research efforts have focused on predicting the expected spreading behavior of contact-based infectious diseases, exploiting characteristics of the population and the disease itself. However, there have been limited research efforts focusing on the use of future social network data: whereas current social network models are abstract constructs where people are anonymously represented, it is not unreasonable to expect developments in data collection (through Facebook, Twitter, Foursquare, etc.) that will allow accurate mappings between known individuals. Furthermore, spatial analysis of transport and communication networks that exploits these data sources is a growing area of research (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) and has recently been expanded to include social network modeling, specifically, the ability to reproduce spatial structure and interaction between individuals for large-scale social networks (Illenberger et al. 2013). The ongoing development of activity-based travel models, which examine why, where, and when various activities are engaged in by individuals (Lam and Huang 2003; Roorda et al. 2009; Ramadurai and Ukkusuri 2010; Illenberger et al. 2013), as well as innovations in pedestrian modeling (Hoogendoorn and Bovy 2005) present additional promising alternatives to generate social-contact networks in the future. Of significant interest is the public transit system, which represents a potential catalyst in the disease transmission process within metropolitan regions. Advances in public transit modeling 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) (Nassir et al. 2012; Pendyala et al. 2012). Although these methods will potentially allow accurate mappings among known individuals, they do not provide a means to use the data for tracking disease transmission. As such, it is critical to develop methods that can exploit these data as they become available.

The objective of the model proposed in this work is to infer the spatiotemporal path of infection through a social contact network for an ongoing outbreak scenario. This work specifically considers contact-based diseases, which refer to the family of infectious diseases that are transmitted from an infected to susceptible individual through direct contact. This category includes sexually transmitted diseases, various strands of the flu, SARS, and the common cold, among others. In turn, the social contact network is representative of the social interactions (e.g., through school, work, or home) that occur among a group of individuals in a given time period (e.g., a day).

Specifically, the problem approached in this paper considers the case in which the structure of the network is deterministically known (set of nodes and links), time-of-infection data are available for all infected nodes, but no information is known about the infection tree (i.e., the set of social contacts through which the disease spread). The problem is formulated as a linear integer program.

[1]School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, NSW 2052, Australia (corresponding author). E-mail: l.gardner@unsw.edu.au

[2]School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, NSW 2052, Australia. E-mail: d.fajardo@unsw.edu.au

[3]Professor, School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, NSW 2052, Australia. E-mail: s.waller@unsw.edu.au

Note. This manuscript was submitted on November 28, 2012; approved on December 11, 2013; published online on May 22, 2014. Discussion period open until October 22, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Natural Hazards Review***, © ASCE, ISSN 1527-6988/04014004(10)/$25.00.

Instead of solving the integer program directly, a much more efficient solution method is developed that exploits properties of the problem. Specifically, a maximum-likelihood estimation procedure is implemented on a simplified acyclic network to efficiently solve the problem. The proposed methodology makes use of network-based optimization algorithms, network structure, and disease properties to determine infection paths based on time-dependent infection reports.

The novelty of the model arises from the utilization of optimization methods (rather than enumeration followed by a posteriori analysis) to infer a contagion process through a network. The model performance is based on how accurately it predicts the paths of infection (which are extracted from simulation outputs). Furthermore, the model's performance is shown to vary significantly as a function of network structure and transmission probability, it but performs best for heterogeneous networks, which are consistent with many real world systems.

The proposed model specifically aims to

1. Provide a novel method for evaluating a region that has been exposed to infection;
2. Provide insight into future epidemic outbreak behavior; and
3. Aid in the evaluation and recommendation of intervention strategies.

In the application of the model evaluated in this paper, individuals are modeled explicitly, and social contacts define the network structure. In previous work by Gardner et al. (2012), similar methodology was implemented to infer the most likely air travel routes responsible for spreading the Swine Flu to unexposed geographic regions. In Gardner's paper, social contact networks were not considered, and the network structure was defined by the air traffic system. In addition, the proposed solution methodology can be extended to alternative contagion processes that occur atop known network structures (e.g., tracking food-borne outbreaks that propagate though a distribution network or computer viruses that spread through a communications network). The remainder of this manuscript includes a literature review of related epidemiology models, problem definition, mathematical formulation and solution methodology, followed by numerical results and conclusions.

Literature Review

Dynamic contagion processes impact copious network systems and are therefore the focus of various studies within the emerging field of network science. In addition to the transmission of infectious disease through communities and biological systems (Murray 2002; Anderson and May 1991), the spread of information, ideas, and opinions through social networks can also be modeled as a contagion process (Coleman et al. 1966; Hasan and Ukkusuri 2013) as well as the global spread of computer viruses on the Internet (Newman et al. 2002; Balthrop et al. 2004), power grid failures in electricity markets (Kinney et al. 2005; Sachtjen et al. 2000), and the collapse of financial systems (Sornette 2003). Of interest to this study is the propagation of disease through a social contact network, which therefore will be the focus of the literature review.

The infection rate and pattern of the disease-spreading process through a network is dependent on both the parameters of the disease (e.g., infectious period, level of contagiousness) and the fundamental structure of the network. In efforts to predict expected disease-spreading behavior and characteristics, epidemiological models span from extremely generalized and simplified analytical models to increasingly in-depth stochastic agent-based simulation tools. Analytical models are used to quantify the statistical properties of epidemic patterns (Colizza et al. 2006; Balcan et al. 2009); however, they are unable to capture certain behavioral aspects of the dynamics of disease spreading and often lack detailed information about the network structure. In contrast, agent-based simulation models can be used to replicate possible spreading scenarios, predict average spreading behavior, and analyze various intervention strategies for a given network and disease while capturing a greater degree of detail, but in turn require a highly detailed set of input data (Rvachev and Longini 1985; Epstein et al. 2002; Eubank et al. 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 et al. 2006; Germann et al. 2006; Ekici et al. 2008; Roche et al. 2011; Haydon et al. 2003). The most recent and comprehensive models provide a greater degree of realism but are difficult to implement within the short time frames in which real-time control decisions must be made. Large-scale simulation models can also be computationally taxing, because multiple runs are required to accurately predict expected outcomes.

There currently exists a gap in the literature that calls for scenario-specific disease prediction models. Most contagion models predict future potential outbreak scenarios based on system-wide information; however, they are not able to reconstruct the contagion process of an ongoing outbreak to reveal information about the current state of the network. Recent advances in disease modeling have begun addressing this issue. For example, there are models that use genetic sequencing data to analytically infer the geographic history of a given virus's migration (Drummond and Rambaut 2007; Lemey et al. 2009; Wallace et al. 2007; Cottam et al. 2008; Haydon et al. 2003). Often this approach involves first enumerating all possible evolutionary trees, then assigning posterior probabilities based on specifics of the respective virus' mutation rates. Additionally, the infection trees only include locations where samples were available. Jombart et al. (2009) proposed a novel approach to reconstruct the spatiotemporal dynamics of outbreaks from sequence data by inferring ancestries directly among strains of an outbreak using their genotype and collection date. The "infectious" links were selected such that the number of mutations between nodes is minimized. The idea of using infection data to construct the most likely path of transmission is the highlighted goal of this paper. The proposed approach relies instead on available infection reports, contact network structure, and disease properties to infer the spatiotemporal path of infection through a contact network.

The set of infection-spreading links identified by the model can provide insight into the spreading behavior of a disease because the contact types most likely to have spread infection are revealed. The expected role that each type of contact plays in the disease-spreading process can then be quantified. The methodology can also be used to evaluate various intervention strategies by adjusting properties of specific link types (e.g., removing or reducing the transmission probability of certain school links to represent a school closure) and comparing the resultant likelihood of a given outbreak scenario. However, to implement this type of analysis, a critical level of detail on the network structure is required, which is just in the initial stages of becoming available.

Problem Definition

Using infection reports, contact network structure, and disease properties, the methodology described in this section makes inferences about infection-spreading patterns in a population. The problem assumes an underlying contagion process that can be

represented on a network by a discrete-time, stochastic process. The following terminology is used for the remainder of this paper:

1. t_i, time stamps: the time period at which a node was reportedly infected, or predicted to be infected;
2. p_{ij}, link transmission probability: the probability that an infected node i will infect a susceptible (and adjacent) node j in a single time step; and
3. λ, infectious period: the number of time steps an infected node remains infectious (i.e., is able to infect others) following its own infection; λ can also represent the amount of time before recovery, hospitalization, or some other type of removal from the network.

The problem objective is defined as follows: For a given social contact network that has been exposed to infection [Fig. 1(a)] and known time stamps for all infected nodes [Fig. 1(b)], the authors seek to identify the infection pattern (i.e., a set of contact links) most likely to have produced the known (node level) infection data [Fig. 1(c)].

The network $G \in (V, A)$ is formally defined by a set of nodes, V, which represent a population of individuals; and links, A, which represent physical daily contacts between individuals; N represents the set of individuals that became infected during the time period when population V was exposed to infection; I represents the set of information nodes: a subset of the infected individuals N, which were identified as infected (i.e., they visited a doctor, hospital, pharmacy, etc.). The full information case is the focus of this paper, and defined as follows: The complete set of infected nodes and the time stamp, t_i, for each infected node is available, i.e., $I = N$.

The relationship between the underlying contagion process and the mathematical programming formulation are of specific interest in regard to the problem definition. The link-based infection process, the building block of the network-based contagion process, is introduced first.

Link-Based Infection Process

The link-based infection process consists of a set of link trials that are the basic building blocks of the network-level contagion process. In other words, a given infection scenario at the network level is the result of many individual link-based trials. Each link trial consists of the following evaluation: At a discrete time step t, assume node i is in an infectious state, node j is in a susceptible state, and the two nodes are connected by link (i, j) with a link transmission probability p_{ij}. A successful link trial is defined as when node i infects node j in time step t, and occurs with probability p_{ij}. The probability a link trial is unsuccessful is therefore $(1 - p_{ij})$. A simulation time step t is representative of the latent period, or the amount of time between when an individual contracts the disease and becomes infectious.

The timestamp of node i, t_i, represents the time (e.g., day) at which individual i was infected. The value Δt_{ij} is the time difference between when nodes i and j were infected. If timestamps are known for two directly connected nodes, where node i was infected before node j, the probability that infection occurred between these two nodes can be calculated using Eq. (1):

$$\alpha_{ij} = (1 - p_{ij})^{(\Delta t_{ij} - 1)} p_{ij} \tag{1}$$

The first term represents the probability of one or more "infection delays," and the second term represents the probability of a single successful link trial. Similarly, the probability that the link-based infection of (i, j) will never be successful is defined as γ_{ij}: over Δt_{ij} consecutive link trials, all will be unsuccessful. The value of γ_{ij} is given by Eq. (2):

$$\gamma_{ij} = (1 - p_{ij})^{\min\{\Delta t_{ij}, \lambda\}} \tag{2}$$

The expression is based on the assumption that node i has a limited infectious period and remains infectious for λ time units. This assumption may not be valid for certain diseases (e.g., for some sexually transmitted diseases individuals remain infectious indefinitely). However, there are many diseases for which the assumption of a limited infectious period is valid. It is assumed that λ is a homogeneous parameter for all nodes.

Now, consider a link selection variable x_{ij}, which is equal to 1 if link (i, j) is selected to be part of the infection tree, and 0 otherwise. The probability associated with including link (i, j) is equivalent to the probability of infection having occurred between nodes i and j, and can then be expressed in terms of Eqs. (1) and (2):

$$\pi_{ij}(x_{ij}) = (1 - p_{ij})^{x_{ij}(\Delta t_{ij} - 1)} p_{ij}^{x_{ij}} \gamma_{ij}^{(1 - x_{ij})} \quad \forall\ (i, j) \in A \tag{3}$$

When $x_{ij} = 0$, the link is not included in the infection tree and $\pi_{ij} = \gamma_{ij}$. When $x_{ij} = 1$, the link is included in the infection tree

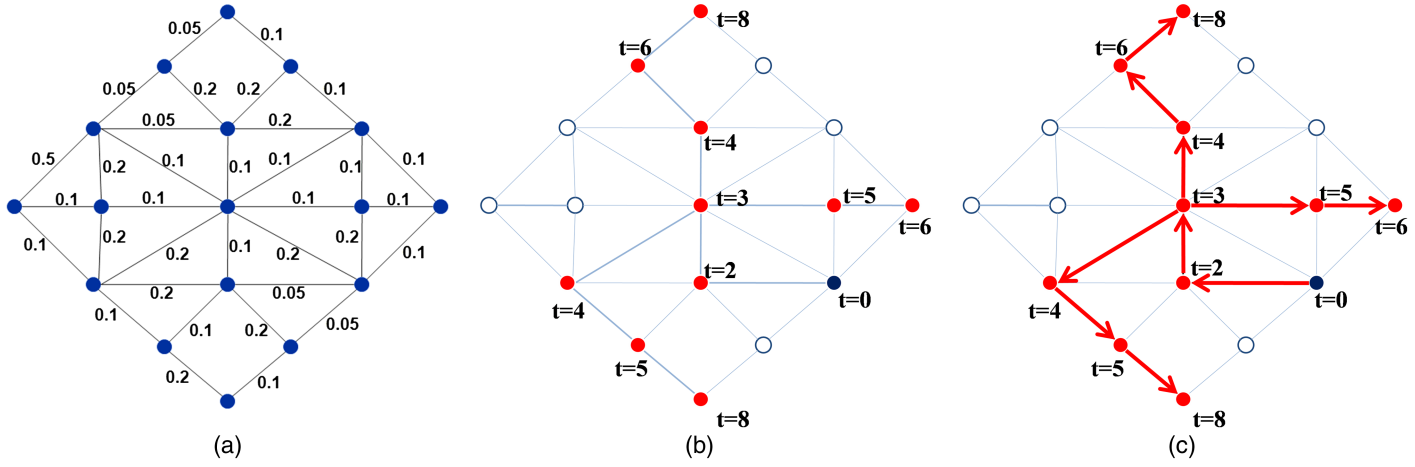

Fig. 1. (a) Sample contact network structure $G \in (V, A)$, and link transmission probabilities, p_{ij}; (b) example of the node level information provided (after an outbreak); set of information nodes, I are solid with associated timestamps, t_i; (c) example of model output (arrows) predicting the infection pattern; the node with timestamp, $t = 0$, is the source

Nat. Hazards Rev., 2014, 15(3): 04014004

and $\pi_{ij} = \alpha_{ij}$. In the next section, this result is extended to the network level.

Network-Based Infection Process

This work treats the network-based infection process as an iterative aggregation of individual link trials. The simulation model starts by initializing all nodes to a susceptible state and randomly choosing a set of nodes to be infected ($O \in V$). Then, transmission of the disease is simulated over multiple time steps, t, for a predetermined simulation period, T. During each time step, all links that connect infectious nodes and susceptible nodes are identified, and infection trials for each such link are performed. If the link infection is successful, then the newly infected node status is changed to "infected" in the following time step. The node remains infected for λ time steps. After a node is infected for λ time steps, its status is changed to "recovered." Once a node is recovered, it can no long transmit the disease or become infected again (the equivalent of gaining immunity or being removed from the network).

This process is representative of a discrete-time network susceptible-infected-removed (SIR) contagion process. The SIR model is a well-established stochastic simulation model used in the epidemiological literature to model the progress of an epidemic in a large population (Grassly and Fraser 2008). The simulation model described previously forms the basis for the mathematical formulation and evaluation presented in the remainder of this paper. It follows that the aim of this solution methodology is to replicate the actual infection tree for a specific outbreak scenario by exploiting node-level infection information and the network structure.

Assumptions

Multiple simplifying assumptions are necessary to solve the proposed problem. This work assumes:

1. A priori knowledge of the underlying social contact network, $G \in (V, A)$;
2. The contagion process can be approximated as discrete-time network SIR contagion process with known transmission probabilities, p_{ij};
3. An individual can be infected at most once, and thus only those diseases for which immunity is acquired after recovery are considered;
4. Known timestamps t_i for the full set of infected nodes, N; and
5. Incubation period, λ is a homogeneous parameter.

The first assumption is the most debatable of the five. Social contact networks are difficult to characterize, as they are not directly observable, and are also highly unstable. However, the increase of social networking information available online and improvements in activity-based travel modeling both contribute towards the possibility of access to more detailed social contact information in the future. Potential sources and methods for generating social contact networks are discussed and cited in the introduction. Assumption two addresses the use of the SIR model, which is used to simulate outbreak patterns in this paper, and a well-established stochastic simulation model (Grassly and Fraser 2008). This simulation tool is therefore the best option to quantify the performance of the proposed model. Quantifying the transmission probability is beyond this scope of this work, but remains a highly researched topic among epidemiologists. Various methods exist to quantify this parameter for known diseases; therefore, the transmission probability is assumed to be available from prior clinical and epidemiological studies for the disease in question. The third assumption restricts the set of applications to those diseases for which acquiring immunity restricts an individual from being infected more than once over the entire course of an outbreak. Assumption four is based on the premise that information is made available for all infected individuals (i.e., through some type of medical authority hospital, private clinic, pharmacy). Although the assumption of full infection data may seem limiting at the present time, various online global health databases and real-time reporting currently exist through organizations such as the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), Centers for Infectious Disease Research and Policies (CIDRAP), and the International Society for Infectious Diseases (through ProMED), among others. These databases are becoming increasingly relied on for tracking outbreaks. Assumption five requires an average incubation period to be applied homogenously across all individuals.

Method

In this section, the mathematical formulation and solution methodology for the full information case are presented. First, the partial information case is briefly discussed, and a linear programming formulation for the full information case is provided. The solution method for the full information case is then described.

Mathematical Problem Formulation

The linear integer programming formulation and proposed solution methodology for the full information case follows:

$$\max_x f(x) = \prod_{\forall (i,j) \in A} (1 - p_{ij})^{x_{ij}(\Delta t_{ij} - 1)} p_{ij}^{x_{ij}} \gamma_{ij}^{(1 - x_{ij})} \tag{4}$$

s.t.

$$t_{\max}(x_{ij} - 1) < t_j - t_i \quad \forall\ (i,j) \in A \tag{5}$$

$$t_{\max}(1 - x_{ij}) + \lambda \geq t_j - t_i \quad \forall\ (i,j) \in A \tag{6}$$

$$\sum_{j \in N} x_{ji} = 1 \quad \forall\ i \in N \setminus O \tag{7}$$

$$x_{ij} = \{0, 1\} \quad \forall\ (i,j) \in A \tag{8}$$

The decision variable is x_{ij}, which is set to 1 if link $(i \cdot j)$ is included in the infection tree and 0 otherwise. The objective [Eq. (4)] enforces that the set of links included in the final spanning tree maximizes the likelihood of the tree; the product of the individual probabilities corresponding to each link included in the tree, as defined in Eq. (3). The first two constraints enforce consistency between the x variable and the contagion process: if $x_{ij} = 1$, meaning i is the predecessor of j in the infection tree, then Eq. (5) guarantees that the infection time of j will be later than that of i. Eq. (6) ensures that that the infection period of j will be within the time period when i is infectious, i.e., within λ time units of the infection time of i. The variable $t_{\max}$ is set equal to the largest time stamp, and representative of M in the *Big M method* traditionally used for solving linear programming problems. Eq. (7) enforces the spanning tree structure of the solution; every known infected node, except the source node, is infected exactly once (i.e., has exactly one incoming link). Eq. (8) declares the decision variable x_{ij} to be binary.

Link Cost Transformation

The objective function [Eq. (4)] can be transformed from a product of terms to an equivalent summation of terms by maximizing the natural logarithm of the objective function, which results in the following:

$$\max_x \sum_{\forall (i,j)\in A} x_{ij}[(\Delta t - 1)\ln(1-p_{ij}) + \ln p_{ij} - \ln\gamma] + \ln\gamma_{ij} \quad (9)$$

The new formulation features additive terms rather than multiplicative terms. The set of constraints remains the same with the addition of (*11*) the new link cost function definition:

$$\max_x f'(x) = \sum_{\forall (i,j)\in A} x_{ij}P_{ij} \quad (10)$$

s.t.

$$P_{ij} = (\Delta t_{ij} - 1)\ln(1-p_{ij}) + \ln p_{ij} - \ln\gamma_{ij} \quad \forall\ (i,j)\in A \quad (11)$$

$$t_{\max}(x_{ij} - 1) < t_j - t_i \quad \forall\ (i,j)\in A \quad (12)$$

$$t_{\max}(1 - x_{ij}) + \lambda \geq t_j - t_i \quad \forall\ (i,j)\in A \quad (13)$$

$$\sum_{j\in N} x_{ji} = 1 \quad \forall\ i \in N \setminus O \quad (14)$$

$$x_{ij} = \{0,1\} \quad \forall\ (i,j)\in A \quad (15)$$

Solution Methodology

Eqs. (10)–(15) allows one to exploit specific properties to develop a much more efficient solution method than solving this linear program directly. As stated previously, this work seeks the most likely infection-spreading pattern for the full information case. The properties of the full information case result in a spanning tree that branches to every node $i \in N$. The general problem of finding a directed maximum branching tree can be solved using the algorithm developed by Edmonds (1967). The algorithm consists of maintaining an optimal subnetwork that reaches every node and works towards feasibility by replacing links that form a cycle in that subnetwork. As such, the cycle-finding subroutine of the algorithm is the most computationally taxing part of the algorithm.

Although Edmonds' algorithm is relatively efficient, a significantly more efficient algorithm is developed here by exploiting the problem properties already presented in this section. Attention is first focused on Eqs. (12) and (13), which allow for the efficient pruning of the set of links that must be considered as part of the procedure. The resulting network is acyclic, which greatly simplifies the maximum branching procedure.

Let the set of feasible links $(i,j)\,\epsilon\, L$ be such that $t_i < t_j < (t_i + \lambda)$. It is trivial to show that in all feasible solutions, $x_{kl} = 0$ for all links (k,l) in $A \setminus L$. Therefore, focus can be limited to the link set L. Because of Eq. (12), which requires feasible links in L to connect nodes with increasing timestamps, the resulting subnetwork has a topological ordering, and as such, cannot contain any directed cycles. Eqs. (14) and (15) represent the requirement that exactly one incoming link is chosen for every infected node. As such, any solution that only chooses links in L so that every node has exactly one incoming link will be feasible. The mathematical program can be written as follows:

$$\max_x f'(x) = \sum_{\forall (i,j)\in S} x_{ij}P_{ij} \quad (16)$$

s.t.

$$P_{ij} = (\Delta t - 1)\ln(1-p_{ij}) + \ln p_{ij} - \ln\gamma_{ij} \quad \forall\ (i,j)\in L \quad (17)$$

$$\sum_{j\in N} x_{ji} = 1 \quad \forall\ i \in N \setminus O \quad (18)$$

$$x_{ij} = \{0,1\} \quad \forall\ (i,j)\in L \quad (19)$$

The most computationally intensive portion of Edmonds' maximum branching algorithm is the search for and removal of cycles. Because of the acyclic nature of the set L, a simplified version of Edmonds' algorithm can be developed that is significantly more efficient, defined by steps 1–3:

1. Define the set of feasible links, $L = \{(i,j): t_i < t_j < (t_i + \lambda)\}$;
2. Calculate link costs, P_{ij} for links (i,j) in feasible set L using Eq. (17); and
3. For each infected node, $j\epsilon N$, select the incoming link (i,j) with the highest cost, P_{ij}, from the set of feasible links L, and add it to the solution tree S.

The set of selected links S forms the maximum likelihood tree.

Network Structures

The network structures evaluated in this work are intended to represent social contact networks. Two types of networks are constructed:

1. Randomly generated: The structure of the network (number of contacts per individual) is determined by a prespecified degree distribution; and
2. Activity-based: The network structure is generated using regional demographic characteristics and human activity patterns.

For a given network, links may have homogenous transmission probabilities, p, or heterogeneous transmission probabilities, in which case, each link is assigned a transmission probability, p_{ij}. This transmission probability can be a dependent on the type of interaction between the two individuals i and j, (e.g., school, work, social). The actual transmission probability values are dependent on the specific characteristics of the disease in question and are not the focus of this study. Therefore, the values selected for the analysis are not specific to any particular virus, and a analysis is conducted to evaluate the sensitivity of the model performance to changes in these parameters.

Urban Network

The urban network generated is representative of a social contact network for a community of individuals that interact on a daily basis though activities such as school and work. In the future, this information could be made available by activity-based travel models or online social network data. However, the sample network analyzed in this paper was created using a synthetic data set consistent with the demographic characteristics of Travis County, Texas, taken from the 2008 U.S. Census Bureau (2008).

The urban network has multiple link types, dependent on the type of trip-based contact (e.g., school, work, social) and corresponding heterogeneous link properties (activity-based transmission probabilities). For example, a node (representing child A) might have two school links (connecting child A to child B and child C) representing contacts at school, and a social link

representing contact with a neighbor they interact with after school. A social link and school link can have different link transmission probabilities. The network structure is defined by the set of nodes and the complete set of link types. For a set number of nodes the number of links in the network will vary as a function of the level of connectivity specified when generating the network. To generate the network, the following steps were taken:

1. Randomly assign an age to each node selected from a given age distribution;
2. Assign all individuals to households either of size 1, 2, 3, or 4;
3. Assign all children (individuals under 18) to a school;
4. Assign all adults (individuals 18-65) to a place of work;
5. Create link connections:
   a. Connect all individuals who share a household with probability 1 (if two individuals share a home link, then they don't share any other links);
   b. Connect all children at the same school to each other with a probability 0.2;
   c. Connect all adults assigned to the same work office with probability 0.1;
   d. Create random shopping links between any two nodes with probability 0.01;
   e. Create random social connections between any two nodes with probability 0.005; and
6. Assign link probabilities $p()$, dependent on link type.

The generated network has 250 nodes and 979 links. The transmission probabilities $p()$ and link connectivity probabilities used for the base case are listed in Table 1. The urban network created is a relatively homogenous network structure; the majority of nodes have degree values close to the network average. The network degree distribution is most similar to a Poisson distribution, with an average degree approaching nine. To contrast this type of network structure, the model performance is compared with various power law network structures.

The probabilities used to connect this network and the corresponding transmission probabilities are not representative of any specific virus. Because the networks presented here are intended to evaluate the proposed methodology, region-specific structural inaccuracies are not yet of vital importance. That being said, developing accurate input data remains a valuable future research problem, specifically (1) calibrated transmission probabilities based on historical outbreak data, and (2) social contact networks extracted from activity-based travel models.

Power Law Networks

Power law networks are the most common structures used to represent social contact networks, which warrants their inclusion in this analysis (González et al. 2008). Power law networks have a degree distribution $f(x) = ax^k$. In this analysis, the exponent parameter k is in the range [1, 3]. The higher exponent k corresponds to a more heterogeneous degree distribution. As k approaches one, the network structure begins to display more uniform degree characteristics. In addition, the number of links increases significantly as k decreases, for the same number of nodes. This is because (for the same number of nodes) there are more nodes with a higher number of connections. The networks are generated according to the method developed by Viger and Latapy (2005). For the power law networks generated, the transmission probabilities are assumed to be homogeneous and are analyzed across a range of values, (0, 1).

Table 1. Urban Network Parameters for Base Case

Parameter	Value
Number of links	979
Probabilities used to create random network links	
Work	0.1
Shopping	0.01
Social	0.005
School	0.2
Probabilities of transmission used in simulation	
Home	0.2
Work	0.1
Shopping	0.05
Social	0.1
School	0.1

Measure of Performance

The proposed methodology is likely to perform differently depending on the network structure and properties of the disease. Therefore, sensitivity analysis was conducted to compare the performance across various combinations of network structures (urban, power law), network sizes (in terms of number of nodes), and disease parameters (transmission probabilities). Although the solution method itself does not require the use of a microscopic level stochastic simulation model for implementation, a simulation model capable of replicating the progress of an epidemic in a large population was used in to evaluate the performance of the proposed methodology. The model performance is measured by comparing the set of links identified in the simulation-based scenario, K (e.g., the actual set of infection spreading links), with those identified by the model, S. The following steps were used to evaluate the performance of this solution methodology:

For a given network structure, $G \in (V, A)$, with known link transmission probabilities, p_{ij}.

1. Set the infectious period λ and time period, T;
2. Randomly introduce an infected individual into the network, O;
3. Simulate an infection spreading scenario for the time period, T;
4. Extract the Full set of links in the infection tree, K from the simulation to use for evaluating the solution methodology;
5. Extract the following information from the simulation to use as (required) input for the solution methodology:
   a. Full set of infected nodes, N;
   b. Timestamps for each infected node, $t_i \forall i \in N$;
6. Implement the solution algorithm (steps 1–3 presented in the solution methodology section) on the extracted network $G \in (N, L)$;
7. Compute the percentage of correctly predicted links, q, for the simulated contagion scenario:
   a. Identify the set of links $M \in K$, where M is the set of links in the output tree S;
   b. $q = |M|/|K|$, that is the percentage of infection links correctly identified by the model; and
8. Repeat steps (1)–(7) X times and average q (step 7.b) over all iterations.

The preceding outlined procedure returns the *expected* performance of the solution methodology, Q, which is how accurately S represents the actual spreading scenario on average, for a given network. For both network structures, Q is based on $X = 1{,}000$ iterations. This analysis is performed for various combinations of network structures, sizes, and disease parameters. The results are presented in the following section.

Numerical Results and Analysis

The expected performance of the solution methodology, Q is illustrated in Figs. 2 and 3 for the following network structures, respectively:

1. Urban network
2. Power law network

The urban network used in this analysis has 250 nodes, 979 links, and the original transmission probabilities as defined in Table 1, varying between [05, 0.2]. The power law network also has 250 nodes, although with fewer links because of the heavier tailed distribution (relative to the Poisson-characterized urban network). The power law network used in this analysis has 379 links, the exponent k is 3, and the transmission probability is set at a constant value ($p = 0.5$) for all links

In the figures, each series represents Q for a constant infectious period, λ. The results illustrate a decrease in expected performance, Q, as the simulation time, T increases. The value of Q also decreases as the infectious period, λ, increases for a constant simulation period, T. These are intuitive results because of the stochastic nature of the contagion process, for which the feasible solution set increases with higher infectious periods and simulation times.

The same general behavior is observed for both the urban and power law network structures, although Q performs better for the power law network—remaining above 85% for all cases analyzed, and above 90% for $T < 9$. In the urban network, Q remains above 80% for $T < 7$, but falls to 60% for higher (λ, T) combinations. The difference in performance among the networks can be partially attributed to the network structures, specifically the high average connectivity in the urban network relative to the power law network. Power law networks have a hub and spoke structure, with a small percentage of highly connected nodes (known as super spreaders), with the majority of nodes having a very low degree (one or two). In the case where the initially infected node is selected randomly, a super spreader node in a power law network structure has a low probability of being selected. In addition, the remaining less-connected nodes have limited opportunity to spread infection, therefore the disease is less likely to spread to a significant portion of the population. In more homogenous network structures, such as Poisson networks, most nodes have close to the average degree, resulting in a more connected network. In a more connected network a higher number of infections are likely to occur. The challenge in predicting the set of infection spreading links (the objective of this study) is that a given set of infected nodes (specific to a given contagion scenario) may correspond to multiple feasible (link-level) infection patterns. Furthermore, for the same number of infected nodes, a more connected network structure will have more feasible infection spreading patterns, thus the likelihood of identifying the *actual* infection tree is reduced. This reasoning explains why the model performs better for the power law network compared with the more homogenous urban network structure which has an average degree distribution around 8. The same reasoning explains the improved performance under shorter infectious periods and shorter simulation times, for which the number of feasible infection trees is minimized.

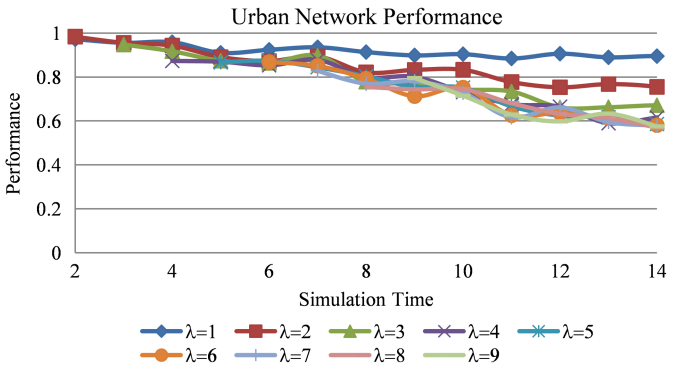

Fig. 2. Expected performance, Q, for urban network

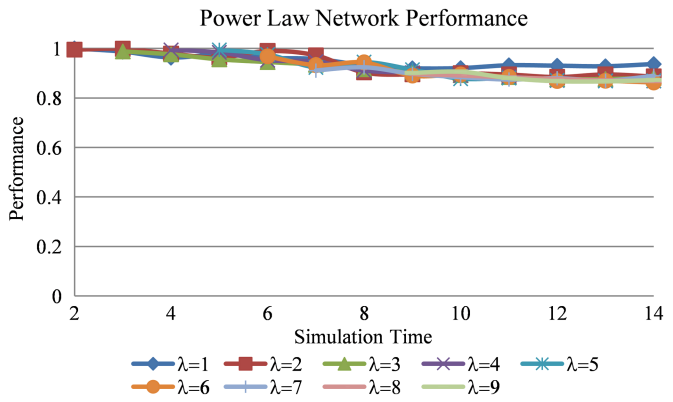

Fig. 3. Expected performance, Q, for power law network

Sensitivity to Transmission Probability

As previously discussed, accurately quantifying the transmission probability for a given disease is beyond the scope of this work. However sensitivity analysis was conducted for urban and power law network structures for varying transmission probabilities to explore the model performance for different potential diseases. In the analysis the infectious period and simulation time were kept constant at $\lambda = 3$ and $T = 10$, respectively.

In Fig. 4, Q is illustrated for the urban network subject to varying transmission levels (which increase along the x-axis). The original transmission probabilities are inflated and deflated by a constant factor, thus remaining proportional to the original activity-specific values. The maximum inflation factor is five, at which point some of the links take on a transmission probability value of one. The maximum deflation is 0.01, which results in many p_{ij} values close to zero. These inflation and deflation factors are chosen such that $(0 < p_{ij} < 1)$. The results are again averaged over 1,000 iterations.

If all the transmission probabilities were zero or one, then the exact scenario would be predicted for S because the infection process would actually be a deterministic process, (with the exception of arbitrarily broken ties, when two different nodes could have resulted in the infection of a third). However, because the transmission probabilities vary across the links, this scenario is not possible to replicate for the urban network. However, the figure shows Q performs very well for extremely low transmission probabilities (when the transmission probabilities are deflated to near zero) due to the decreased level of uncertainty associated with the infection process. Under this condition, the infection process is nearly deterministic. The lower performance when the transmission probabilities are midrange is a function of the extreme variation in transmission probability values across links, which range from. 05 to 1, and the increased level of uncertainty in the contagion process.

In Fig. 5 a similar analysis is presented for a power law network, specifically exploring the robustness of Q to varying levels of transmission probability. The power law network generated for

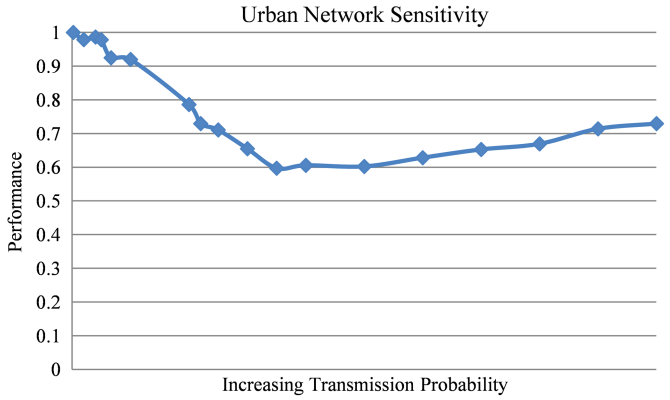

Fig. 4. Urban network sensitivity to transmission probability

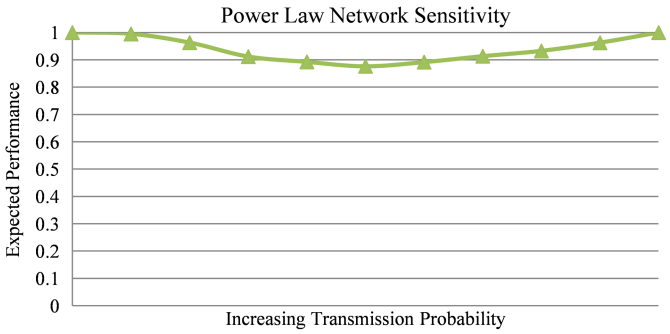

Fig. 5. Power law network sensitivity to transmission probability

this analysis has 1,000 nodes, 1,565 links, and the exponent k is set to 3. Once again $\lambda = 3$, and simulation time, $T = 10$.

Initially (in Fig. 3), the transmission probability was set to $p = 0.5$ for all links. To illustrate the robustness of Q to the transmission probability, Q is presented for 14 discrete values of p, ranging between (0,1) for the power law network. The results are averaged over 200 iterations. In Fig. 5, Q is illustrated across the range of link transmission probabilities for the network.

When the transmission probability is zero or one for all links $Q = 1$, this is representative of a deterministic case that can be computed with 100% accuracy. For the nondeterministic cases $(0 < p < 1)$ the performance varies, with the lowest performance corresponding to $p = 0.5$. This is intuitively the hardest case to predict because there will be multiple infection scenarios with equal probability. However, even for $p = 0.5$, the performance still remains close to 90%. A likely factor in the high performance can be attributed to the highly heterogeneous network structure.

The increasing performance as the transmission probability increases from 0.5 can be explained using the known algorithmic behavior for networks with homogenous transmission probabilities: the link with the smallest infection delay is always chosen as the infection-causing contact. More specifically, when all link transmission values are identical, the link ranking will strictly depend on the Δt_{ij} value. The adjacent node i most recently infected $(\min_i \Delta t_{ij})$ will always be chosen as the predecessor node because $(1-p)^n(p) > (1-p)^m(p), \forall 1 < n < m$. This also means that, for a given set of timestamps and homogenous p value, the algorithm will predict the same spreading scenario. For higher p values, this property is more likely to be accurate (higher transmission probability will more often result in immediate infection). Similarly, the decrease in Q as p increases from the deterministic case, $p = 0$, can be attributed to the fact that a longer infection delay is more likely to occur at lower transmission probabilities, although the link with the smallest delay is always going to be chosen. Therefore, S is less likely to replicate the actual spreading scenario.

Sensitivity to Transmission Probability Accuracy

The final analysis presented explores the sensitivity of Q to the accuracy of the p value assigned. In the situation where a disease's properties are unknown, which is a likely situation at the onset of an outbreak when the proposed methodology is intended to be implemented, estimated values of the transmission probabilities must be used. The previous analysis explored the model's performance subject to known p values. To explore the robustness of the model to the accuracy of p, a p' value is selected which differs from the actual p value by Δp:$p' = p + \Delta p$. This estimated p' is used in the link costs to calculate S; Δp can be positive or negative, as long as $0 < (p + \Delta p) < 1$ and simply represents the inaccuracy of the transmission probability value assumption. For example, when $p = 0.5$ (this is the actual transmission probability that dictates the behavior of the outbreak), and $\Delta p = -0.3$, S is determined using $p' = 0.2$ and not the true value, $p = 0.5$. This represents a case where the disease is thought to be much less contagious than it actually is. The value of Q is still calculated by comparing the actual spreading scenario (here represented using a simulation with a specified p), with the computed scenario, $S_{p'}$ determined using p'. The impact of using p' (instead of p) on the model performance is illustrated by comparing Q_p with $Q_{p'}$. These values are the expected performance under the original assumption that the correct p is known when solving S_p, Q_p, and the expected performance when p' is used to compute $S_{p'}$, $Q_{p'}$, $\Delta Q = (Q_p - Q_{p'})$, which represents the difference in performance under the two information assumptions.

For the networks with a homogenous link transmission probability, the link infection probability, α_{ij}, is strictly decreasing in Δt regardless of the value of p, and the exclusion probability γ is the same for any link that is not chosen. As such, a fixed deviation affecting all links (e.g., all transmission probabilities are over/under valued equally) will not affect the ordering with respect to link costs, and therefore will not alter the inferred spanning tree S; the adjacent node i most recently infected $(\min_i \Delta t_{ij})$ will always be chosen as the predecessor of node j.

If the transmission probabilities are heterogeneous, inaccuracies in the transmission probabilities will affect the ordering of links with respect to their infection probabilities. For a constant Δp across all links $(p'_{ij} = p_{ij} + \Delta p)$, the link infection probability $\alpha'_{ij} = (1 - p'_{ij})^{(\Delta t - 1)}(p'_{ij}) = [1 - (p_{ij} + \Delta p)]^{(\Delta t - 1)}(p_{ij} + \Delta p)$, and the exclusion probability $\gamma'_{ij} = (1 - p'_{ij})^{\min\{\Delta t_{ij}, \lambda\}} = [1 - (p_{ij} + \Delta p)]^{\min\{\Delta t_{ij}, \lambda\}}$. The incoming link selected for a given node j will be that which maximizes the product of all (feasible) incoming link probabilities, $\prod_{i \in A_j} \pi_{ij}(x_{ij})$, where A_j is the feasible adjacency list for node j. The same property holds for the homogenous case; however, with heterogeneous links, the exclusion probabilities are no longer equivalent.

For the following analysis, refer to the simple three-node network illustrated in Fig. 6. By Eq. (3) if $x_{ik} = 1$, $\pi_{ik} = \alpha_{ik}$ and if $x_{ik} = 0$, then $\pi_{ik} = \gamma_{ik}$. The link included in the spanning tree S will be that which maximizes $\pi_{ik}\pi_{jk}$. If link (i, k) is selected as the incoming link, then $\pi_{ik} = \alpha_{ik}$ and $\pi_{jk} = \gamma_{jk}$ and $(\alpha_{ik}\gamma_{jk}) > (\gamma_{ik}\alpha_{jk})$. Otherwise $\pi_{ik} = \gamma_{ik}$ and $\pi_{jk} = \alpha_{jk}$. For heterogeneous link probabilities, if $\alpha_{ik}\gamma_{jk} > \gamma_{ik}\alpha_{jk}$, it is not always true that $\alpha'_{ik}\gamma'_{jk} > \gamma'_{ik}\alpha'_{jk}$. In other words, the inferred

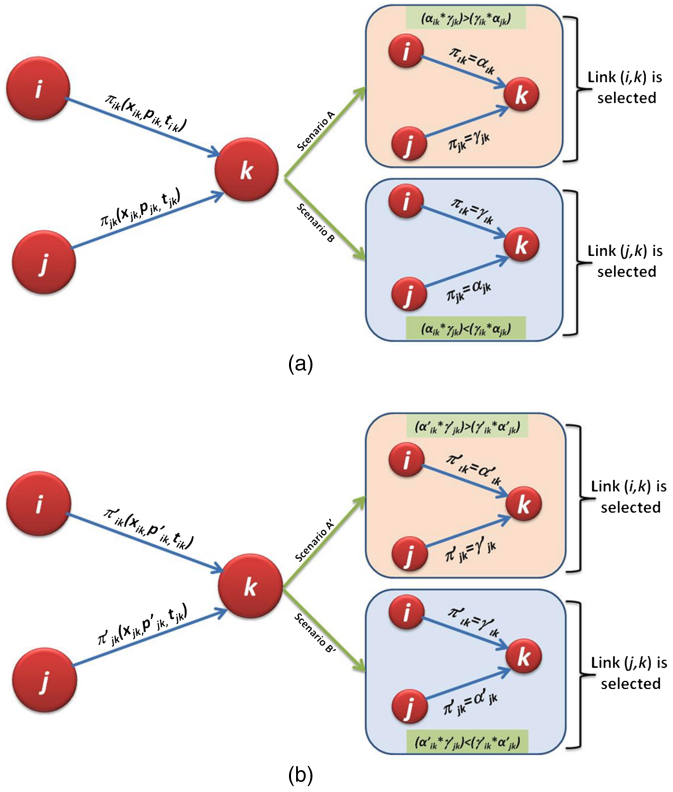

Fig. 6. (a) Example network with link costs and possible infection link selections for a network with accurate transmission probabilities, p; (b) example network with link costs and possible infection link selections for a network with inaccurate transmission probabilities, p'

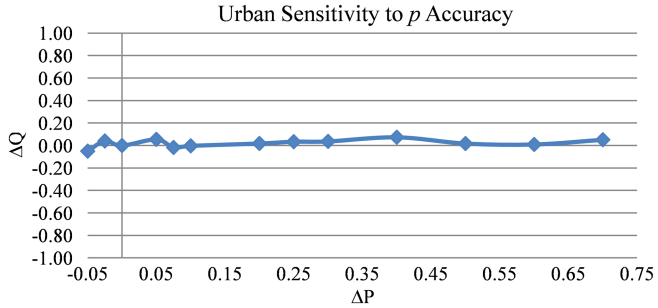

Fig. 7. Urban network sensitivity to accuracy of transmission probability, p

spanning tree may vary (e.g., if scenario A occurs under the actual transmission probability [as in Fig. 6(a)], scenario B' may occur under the estimated transmission probability [as in Fig. 6(b)]. Therefore, the initial ranking of adjacent links will not always remain constant, which will result in a different inferred infection tree ($S_p \neq S_{p'}$).

The model performance for the urban network, with inaccurate heterogeneous link transmission probabilities, is illustrated in Fig. 7. The Δp is chosen such that $0 < p'_{ij} < 1$ for all links, and the original link probabilities are those shown in Table 1. The results are highly robust for Δp values within the allotted range, with a maximum reduction in performance of only 4%. This is an important behavioral property of the model due to the challenge of accurately quantifying transmission probabilities.

Conclusions

The main focus of this study was to develop a formulation and solution method for inferring a contagion process in a social contact network. The proposed methodology provides a novel procedure for evaluating a region that has been exposed to infection (compared with the traditional methods of enumeration followed by a posteriori analysis). The novelty of the model lies in the use of network optimization tools, infection data, and the contact network structure to infer spatiotemporal outbreak patterns, aiding in the development of real-time analysis and decision support for outbreak scenarios. The network optimization model developed was derived through a series of transformations and formulated as a linear program. An extremely efficient solution method was developed to compute the most likely infection spanning tree.

The solution method was tested on two sets of networks, with varying structures, characteristics, and degrees of heterogeneity. In addition, the impact of human error in assessing the disease properties was quantified. Although the performance varied as a function of network structure and transmission probability, the methodology performed best for heterogeneous network structures. This is a favorable outcome because heterogeneous structural properties are characteristic of many real-world networks in which contagion processes occur. In addition, the methodology performance was shown to be robust to estimation errors in terms of transmissibility, another favorable characteristic of the model due to the challenge of accurately estimating disease transmission properties.

The largest weakness with the proposed methodology is the lack of verifiability because of limited data availability. Without social contact network structures and corresponding link-level infection data to calibrate the model, it is not possible to truly evaluate certain model characteristics. Therefore a major motivation for this work is to incentivize specific data collection efforts. For the proposed model, collecting contact-level infection data would be the most valuable. Contact level information requires data from infected individuals on their recent social interactions. Such link-level infection data are difficult to collect, but would permit quantitative analysis of the models' performance. Additionally, research in the development of social networks that more accurately depict the true network structure of a region will be integral in the implementability of this research. Research in activity-based travel modeling can contribute towards defining social contacts. By using regional travel patterns (such as origin-destination tables and activity-based travel patterns), individuals' daily trips, specific types of interaction, and length of interaction can be accounted for, also aiding the transmission probability estimation.

Finally, the proposed model has multiple potential extensions that will be expanded on in future work. One such extension includes the case of partial information. In a more realistic setting, only a fraction of infected individuals will consult a physician, visit a hospital, etc., limiting the availability of information. The objective of the partial information case will be to determine the most likely set of infection spreading contacts when only a subset of the infected nodes are identified, while concurrently identifying the unreported infection nodes. The resulting problem is in fact a nonlinear, nonconvex mixed-integer program, which is computationally intractable. As such, the focus of the research will be to find efficient heuristics to solve the problem.

References

Anderson, R. M., and May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*, Oxford Univ. Press, Oxford, U.K.

Balcan, D., et al. (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases." *Proc. Natl. Acad. Sci.*, 106(51), 21484–21489.

Balthrop, J., Forrest, S., Newman, M. E. J., and Williamson, M. M. (2004). "Technological networks and the spread of computer viruses." *Science*, 304(5670), 527–529.

Cahill, E., Crandall, R., Rude, L., and Sullivan, A. (2005). "Space-time influenza model with demographic, mobility, and vaccine parameters." *Proc., 5th Annual Hawaii Int. Conf. Math., Statist., and Related Fields*.

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G., and Barabási, A. L. (2008). "Uncovering individual and collective human dynamics from mobile phone records." *J. Phys. A: Math. Theor.*, 41(22), 224015.

Carley, K. M., et al. (2006). "BioWar: Scalable agent-based model of bio-attacks." *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans*, 36(2), 252–265.

Coleman, J., Menzel, H., and Katz, E. (1966). *Medical innovations: A diffusion study*, Bobbs Merrill, New York.

Colizza, V., Barrat, A., Barthélemy, M., and Vespignani, A. (2006). "The modeling of global epidemics: Stochastic dynamics and predictability." *Bull. Math. Biol.*, 68(8), 1893–1921.

Cottam, E. M., et al. (2008). "Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus." *Proc. R. Soc. B*, 275(1637), 887–895.

Dibble, C., and Feldman, P. G. (2004). "The geograph 3D computational laboratory: Network and terrain landscapes for repast." *J. Artif. Soc. Social Simul.*, 7(1).

Drummond, A. J., and Rambaut, A. (2007). "Beast: Bayesian evolutionary analysis by sampling trees." *BMC Evol. Biol.*, 7(1), 214.

Dunham, J. B. (2005). "An agent-based spatially explicit epidemiological model in MASON." *J. Artif. Soc. Social Simul.*, 9(1).

Edmonds, J. (1967). "Optimum branchings." *J. Res. Natl. Bur. Stand.*, 71B(4), 233–240.

Ekici, A., Keskinocak, P., and Swann, J. L. (2008). "Pandemic influenza response." *Simulation Conf., 2008 WSC*, 1592–1600.

Epstein, J., et al. (2002). *Toward a containment strategy for smallpox bioterror: An individual-based computational approach*, Brookings Institute Press, 55.

Erath, A., Löchl, M., and Axhausen, K. (2009). "Graph-theoretical analysis of the Swiss road and railway networks over time." *Networks Spat. Econ.*, 9(3), 379–400.

Eubank, S., et al. (2004). "Modeling disease outbreaks in realistic urban social networks." *Nature*, 429(6988), 180–184.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. (2006). "Strategies for mitigating an influenza pandemic." *Nature*, 442(7101), 448–452.

Gardner, Lauren M., Fajardo, D., and Waller, S. T. (2012). "Inferring infection spreading links in an air traffic network." *Transportation Research Record 2300*, Transportation Research Board, Washington, DC, 13–21.

Gastner, M. T., and Newman, M. E. J. (2006). "The spatial structure of networks." *Eur. Phys. J. B*, 49(2), 247–252.

Germann, T. C., Kadau, K., Longini, I. M., and Macken, C. A. (2006). "Mitigation strategies for pandemic influenza in the United States." *Proc. Natl. Acad. Sci.*, 103(15), 5935–5940.

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). "Understanding individual human mobility patterns." *Nature*, 453, 479–482.

González, M. C., Lind, P. G., and Herrmann, H. J. (2006). "System of mobile agents to model social networks." *Phys. Rev. Lett.*, 96(8), 088702.

Grassly, N. C., and Fraser, C. (2008). "Mathematical models of infectious disease transmission." *Nat. Rev. Microbiol.*, 6(6), 477–487.

Hasan, S., and Ukkusuri, S. (2013). "Social contagion process in informal warning networks to understand evacuation timing behavior." *J. Public Health Manage. Pract.*, 19, S68–S69.

Haydon, D. T., et al. (2003). "The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak." *Proc. R. Soc. B*, 270(1511), 121–127.

Hoogendoorn, S. P., and Bovy, P. H. L. (2005). "Pedestrian travel behavior modeling." *Networks Spatial Econ.*, 5(2), 193–216.

Hufnagel, L., Brockmann, D., and Geisel, T. (2004). "Forecast and control of epidemics in a globalized world." *Proc. Natl. Acad. Sci.*, 101(42), 15124–15129.

Illenberger, J., Nagel, K., and Flötteröd, G. (2013). "The role of spatial interaction in social networks." *Networks Spat. Econ.*, 13(3), 255–282.

Jombart, T., Eggo, R. M., Dodd, P., and Balloux, F. (2009). "Spatiotemporal dynamics in the early stages of the 2009 A/H1N1 influenza pandemic." *PLoS Curr. Influenza*.

Kinney, R., Crucitti, P., Albert, R., and Latora, V. (2005). "Modeling cascading failures in the North American power grid." *Eur. Phys. J. B*, 46(1), 101–107.

Lam, W. H. K., and Huang, H. (2003). "Combined activity/travel choice models: Time-dependent and dynamic versions." *Networks Spat. Econ.*, 3(3), 323–347.

Lemey, P., Suchard, M., and Rambaut, A. (2009). "Reconstructing the initial global spread of a human influenza pandemic: A Bayesian spatial-temporal model for the global spread of H1N1pdm." *PLoS Curr. Influenza*.

Meyers, L., Pourbohloul, B., Newman, M. E. J., Skowronski, D., and Brunham, R. (2005). "Network theory and SARS: Predicting outbreak diversity." *J. Theor. Biol.*, 232(1), 71–81.

Murray, J. D. (2002). *Mathematical biology*, 3rd Ed., Springer, New York.

Nassir, N., Khani, A., Hickman, M., and Noh, H. (2012). "An intermodal optimal multi-destination tour algorithm with dynamic travel times." *Transportation Research Record 2283*, Transportation Research Board, Washington, DC, 57–66.

Newman, M. E. J., Forrest, S., and Balthrop, J. (2002). "Email networks and the spread of computer viruses." *Phys. Rev. E*, 66(3), 035101.

Pendyala, R., et al. (2012). "Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation." *Transportation Research Record 2203*, Transportation Research Board, Washington, DC, 19–27.

Ramadurai, G., and Ukkusuri, S. (2010). "Dynamic user equilibrium model for combined activity-travel choices using activity-travel supernetwork representation." *Networks Spat. Econ.*, 10(2), 273–292.

Roche, B., Drake, J., and Rohani, P. (2011). "An agent-based model to study the epidemiological and evolutionary dynamics of Influenza viruses." *BMC Bioinf.*, 12(87), 87.

Roorda, M. J., Carrasco, J. A., and Miller, E. J. (2009). "An integrated model of vehicle transactions, activity scheduling and mode choice." *Transp. Res. Part B*, 43(2), 217–229.

Rvachev, L., and Longini, I. (1985). "A mathematical model for the global spread of influenza." *Math. Biosci.*, 75(1), 3–22.

Sachtjen, M. L., Carreras, B. A., and Lynch, V. E. (2000). "Disturbances in a power transmission system." *Phys. Rev. E*, 61(5), 4877–4882.

Schintler, L., Kulkarni, R., Gorman, S., and Stough, R. (2007). "Using raster-based GIS and graph theory to analyze complex networks." *Networks Spat. Econ.*, 7(4), 301–313.

Small, M., and Tse, C. K. (2005). "Small world and scale free model of transmission of SARS." *Int. J. Bifurcation Chaos Appl. Sci. Eng.*, 15(5), 1745–1755.

Sornette, D. (2003). *Why stock markets crash: Critical events in complex financial systems*, Princeton University Press, Princeton, NJ.

U.S. Census Bureau. (2008). "State & county quickfacts: Travis county, TX." ⟨http://quickfacts.census.gov⟩ (Feb. 5, 2008).

Viger, F., and Latapy, M. (2005). "Efficient and simple generation of random simple connected graphs with prescribed degree sequence." *Proc., 11th Conf. on Computing and Combinatorics (COCOON)*, 440–449.

Wallace, R. G., HoDac, H., Lathrop, R. H., and Fitch, W. M. (2007). "A statistical phylogeography of influenza A H5N1." *Proc. Natl. Acad. Sci.*, 104(11), 4473–4478.

Wang, P., González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2009). "Understanding the spreading patterns of mobile phones viruses." *Science*, 324(5930), 1071–1076.