Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 240 (2006) 400–418
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Predicting epidemics on directed contact networks

Lauren Ancel Meyers[a,b,*], M.E.J. Newman[b,c], Babak Pourbohloul[d,e]

[a]Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA
[b]External Faculty, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[c]Center for the Study of Complex Systems, University of Michigan, Randall Laboratory, 500 E. University Ave., Ann Arbor, MI 48109-1120, USA
[d]Division of Mathematical Modeling, University of British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC, Canada V5Z 4R4
[e]Department of Health Care & Epidemiology, University of British Columbia, Canada

Received 11 May 2005; received in revised form 16 August 2005; accepted 6 October 2005
Available online 21 November 2005

Abstract

Contact network epidemiology is an approach to modeling the spread of infectious diseases that explicitly considers patterns of person-to-person contacts within a community. Contacts can be asymmetric, with a person more likely to infect one of their contacts than to become infected by that contact. This is true for some sexually transmitted diseases that are more easily caught by women than men during heterosexual encounters; and for severe infectious diseases that cause an average person to seek medical attention and thereby potentially infect health care workers (HCWs) who would not, in turn, have an opportunity to infect that average person. Here we use methods from percolation theory to develop a mathematical framework for predicting disease transmission through semi-directed contact networks in which some contacts are undirected—the probability of transmission is symmetric between individuals—and others are directed—transmission is possible only in one direction. We find that the probability of an epidemic and the expected fraction of a population infected during an epidemic can be different in semi-directed networks, in contrast to the routine assumption that these two quantities are equal. We furthermore demonstrate that these methods more accurately predict the vulnerability of HCWs and the efficacy of various hospital-based containment strategies during outbreaks of severe respiratory diseases.

Keywords: Epidemiology; Contact network; Directed graph; Infectious disease; Hospital transmission

1. Introduction

Many infectious diseases spread through direct person-to-person contact. Respiratory-borne diseases like influenza, tuberculosis, meningococcal meningitis and SARS, spread through the exchange of respiratory droplets between people in close physical proximity to each other. Sexually transmitted diseases like HIV, genital herpes, and syphilis spread through intimate sexual contact. Explicit models of the patterns of contact among individuals in a community, *contact network models*, provide a powerful approach for predicting and controlling the spread of such infectious diseases (Longini, 1988; Sattenspiel and Simon, 1988; Morris, 1995; Kretzschmar et al., 1996; Ball et al., 1997; Morris and Kretzschmar, 1997; Ferguson and Garnett, 2000; Hethcote, 2000; Lloyd and May, 2001; Newman, 2002; Sander et al., 2002; Keeling et al., 2003; Meyers et al., 2003; Meyers et al., 2005). This approach has provided insight into the impact of simultaneous sexual partners on HIV transmission (Morris and Kretzschmar, 1997) and effective public health strategies for controlling STDs (Kretzschmar et al., 1996) and mycoplasma pneumonia (Meyers et al., 2003), among others.

*Corresponding author. Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA. Tel.: +1 512 471 4950; fax: +1 512 471 3878.
E-mail addresses: laurenmeyers@mail.utexas.edu (L.A. Meyers), mejn@umich.edu (M.E.J. Newman), babak.pourbohloul@bccdc.ca (B. Pourbohloul).

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.10.004

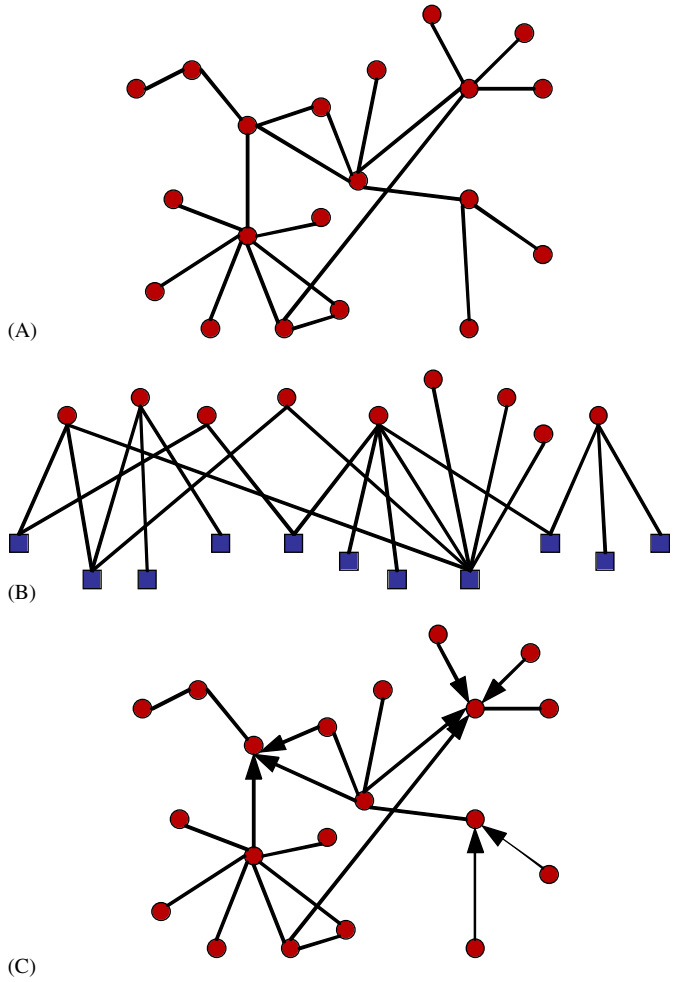

Fig. 1. Contact networks: (A) undirected network; (B) bipartite network; and (C) semi-directed network.

The simplest form of contact network model represents individuals as vertices and contacts as edges connecting appropriate vertices. The undirected network depicted in Fig. 1A assumes that if vertices i and j share an edge, then the probability that i infects j given that i is infective and j is susceptible is equal to the probability that j infects i given that i is susceptible and j is infective. There are many diseases for which this assumption does not hold. For example, there may be as much as a two-fold difference between male-to-female and female-to-male HIV transmission efficiency with females much more vulnerable than males (Nicolosi et al., 1994); health care workers (HCWs) and patients may have asymmetric transmission probabilities because, perhaps, patients are more likely to have immune deficiencies or caregivers are more likely to be exposed to bodily fluids during medical procedures; mothers can transmit blood-borne diseases to offspring in utero whereas there may be no opportunity for transmission in the reverse direction. We can model such asymmetries using bipartite contact networks in which there are two classes of nodes that transmit disease to each other at different rates (Fig. 1B). Mathematical methods for predicting the spread of disease on bipartite contact networks have been described in Ball et al. (1997) and Meyers et al. (2003).

Asymmetry in disease transmission may also arise if the disease influences individual behavior. During an outbreak, infected individuals may modify their typical patterns of interaction. In particular, they may visit a hospital or clinic at which they come into contact with HCWs and other patients. Individuals that are not infected, however, will likely have no contact with hospital personnel. Since we cannot know a priori which individuals will become infected, we cannot easily capture such conditional contacts in a simple network model.

Directed edges, in which transmission occurs only in one direction, provide a way around this difficulty (Fig. 1C). A directed edge leading from a member of the general population (P) to a HCW (H) reflects the following relationship: If P is infected, he or she will expose H with some probability; but if H is infected, he or she will have no contact with P. Thus, contact network models containing both directed and undirected edges (henceforth *semi-directed* networks) can be used to model community-based disease transmission in which there is a substantial one-way flow of disease from the general public into health care facilities. For respiratory diseases, predicting and controlling this flow is vital. Hospitals are

particularly vulnerable because of the frequent hospitalization of infected individuals with serious illness, the high number of patients with pre-existing high acuity co-morbidity including those who are immunocompromised, and the close and multiple physical contacts between infected individuals, caregivers, visitors and other patients. For these reasons, a significant proportion of SARS transmission events occurred within hospitals (Avendano et al.; Varia et al., 2003). Understanding and containing hospital-based transmission is critical not only for the protection of such individuals but also for the prevention of community-wide spread outside the hospital.

Here we develop mathematical tools for predicting the spread of disease and impact of intervention on semi-directed networks and then apply these tools to assess the impact of hospital-based transmission and intervention on the fate of an outbreak. For part one, we use generating function methods to derive the probability and expected demographic distribution of outbreaks, with and without public health intervention. This is an extension of both epidemiological theory previously developed for undirected contact networks (Newman, 2002) and a general theory of random graphs containing only directed edges (Newman et al., 2001). Many of the calculations are fundamentally equivalent to branching process calculations, and it seems likely that some of the results presented here could be derived using branching process methods as well (Jagers, 1975; Andersson, 1998). We show that in semi-directed networks the probability of an epidemic and the expected fraction of the population infected during such an epidemic may be different. In contrast, many conventional models assume the equality of these two epidemiological values, and then use disease incidence data to indirectly estimate the probability of an epidemic (Anderson and May, 1991). Our analysis therefore suggests that this assumption may be invalid for populations with asymmetric contact patterns. For part two, we make epidemiological predictions using a simple model of urban contact patterns based on demographic data from the city of Vancouver, British Columbia. By incorporating conditional contacts within health care settings, we more accurately assess the role of HCWs in disease transmission and containment.

2. Derivations of epidemic quantities

2.1. Modeling the population and the disease

In a semi-directed network, each vertex (individual) has an *undirected degree* representing the number of undirected edges joining the vertex to other vertices as well as both an *in-degree* and an *out-degree* representing the number of directed edges incoming from other individuals and outgoing to other individuals, respectively. The undirected-degree and in-degree indicate how many contacts can spread disease to the individual, and thus is related to the likelihood that an individual will become infected during an epidemic; and the undirected-degree and out-degree indicate how many contacts may be infected by that individual should he or she become infected, and thus is related to the likelihood that an individual will contribute to an epidemic. The *semi-directed degree distribution* tells us the probability that a randomly chosen individual will have a particular combination of an undirected-degree, in-degree, and out-degree.

One can predict analytically the spread of an infectious disease through a population given two basic inputs: the semi-directed degree distribution and the probabilities of disease transmission along the edges of the network. Some pathogens, like smallpox, are highly contagious and will thus have a high probability of moving along an edge in the network (Bozzette et al., 2003). Other pathogens, like SARS, are less likely to be transmitted (Xu et al., 2004). For a given disease, the probability of transmission along a particular edge will also depend on the health of the individuals lying at either end of the edge and the nature of their interaction with each other.

In (Newman, 2002), Newman showed that, when the rate of transmission of a disease between pairs of individuals is assumed to be an i.i.d. random variable, the spread of the disease depends only on the mean total probability of transmission between individuals, or *transmissibility*, and not on the individual probabilities for specific pairs. We make use of this result here also, and henceforth consider only T_d and T_u, the average probability that an infectious individual will transmit the disease to a susceptible individual with whom they have a directed or undirected contact, respectively. Note that average transmissibilities T_d and T_u vary from disease to disease but are always in the range $0 \leqslant T_d, T_u \leqslant 1$. We will also consider the simpler case where average transmissibility is the same for directed and undirected edges, that is, $T_d = T_u = T$.

Suppose a disease begins to spread through a population from a particular vertex. In our model, transmission will occur along each of the directed and undirected edges pointing out of that vertex with probabilities T_d and T_u, respectively. If we keep track of every edge in the network along which disease is transmitted and call these *occupied* edges, then we can reconstruct the final size and distribution of the outbreak. In particular, the outbreak will include exactly the set of all vertices that are connected to the initial vertex along a continuous path of occupied edges. Because of its resemblance to bond percolation, this model can be analysed using mathematical methods from percolation theory (Newman, 2002; Sander et al., 2002; Meyers et al., 2003). In what follows, we derive exact solutions for the expected size of an outbreak, the probability of a large-scale epidemic, the size of such an epidemic, the risk to individuals as a function of their degree, and the impact of various forms of intervention.

2.2. Probability generating functions for semi-directed networks

In the theory of random directed graphs developed by Newman et al. (2001), one considers the joint probability distribution p_{jk} that a randomly chosen vertex has in-degree j and out-degree k. Then one defines a generating function $\mathscr{F}(x,y)$ whose coefficients are the probabilities in this distribution:

$$\mathscr{F}(x,y)=\sum_{jk}p_{jk}x^j y^k \tag{1}$$

from which many properties of the network can then be derived. Adopting a similar approach for our semi-directed networks, we consider the joint probability distribution p_{jkm} that a vertex has j incoming edges, k outgoing edges, and m undirected edges. Then we define a generating function $\mathscr{G}$ that generates this distribution thus:

$$\mathscr{G}(x,y;u)=\sum_{jkm}p_{jkm}x^j y^k u^m. \tag{2}$$

This function has the properties that

$$\mathscr{G}(1,1;1)=1 \tag{3}$$

for any properly normalized p_{jkm}, and

$$z_d=\mathscr{G}^{(1,0;0)}(1,1;1)=\mathscr{G}^{(0,1;0)}(1,1;1),\quad z_u=\mathscr{G}^{(0,0;1)}(1,1;1), \tag{4}$$

where z_d is the average in-degree and out-degree of a vertex for directed edges (the two must necessarily be the same, since every outgoing directed edge must also be an incoming edge at some vertex) and z_u is the average degree of undirected edges. The notation $\mathscr{G}^{(r,s;v)}$ indicates differentiation of $\mathscr{G}$ with respect to its three arguments r, s, and v times, respectively, so that, for example

$$\mathscr{G}^{(0,0;1)}=\frac{\partial\mathscr{G}}{\partial u},\quad \mathscr{G}^{(1,1;0)}=\frac{\partial^2\mathscr{G}}{\partial x\partial y}. \tag{5}$$

The *excess degrees* are the numbers of each type of edge emerging from a vertex arrived at by following an edge, not including the incoming edge. Henceforth, subscript u refers to following an undirected edge in either direction and subscripts d and r refer to following a directed edge in the designated and reverse direction, respectively.

The excess degrees are biased by the fact that edges are more likely to arrive at vertices with higher in- and undirected-degree, in direct proportion to that degree. Thus the distribution of edges of the three types, incoming, outgoing, and undirected, at a vertex reached by following a directed edge in the designated direction is $jp_{jkm}/\Sigma jp_{jkm}$, and hence the excess degree distribution is generated by

$$\mathscr{H}_d(x,y;u)=\frac{\sum_{jkm}jp_{jkm}x^{j-1}y^k u^m}{\sum_{jkm}jp_{jkm}}=\frac{1}{z_d}\mathscr{G}^{(1,0;0)}(x,y;u). \tag{6}$$

The generating function for the excess degree distribution for a directed edge in the reverse direction is

$$\mathscr{H}_r(x,y;u)=\frac{\sum_{jkm}kp_{jkm}x^j y^{k-1}u^m}{\sum_{jkm}kp_{jkm}}=\frac{1}{z_d}\mathscr{G}^{(0,1;0)}(x,y;u). \tag{7}$$

Similarly, the distribution at a vertex reached by following an undirected edge is generated by

$$\mathscr{H}_u(x,y;u)=\frac{\sum_{jkm}mp_{jkm}x^j y^k u^{m-1}}{\sum_{jkm}mp_{jkm}}=\frac{1}{z_u}\mathscr{G}^{(0,0;1)}(x,y;u). \tag{8}$$

We next modify these generating functions to consider the distribution of occupied edges, that is, edges along which disease has been transmitted. In Appendix A.1, we derive the following probability generating function for the number of occupied edges of a vertex:

$$\mathscr{G}(x,y;u;T_d,T_u)=\mathscr{G}(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u). \tag{9}$$

Similarly, the probability generating functions for the excess number of occupied edges, that is, the number of edges (excluding the arrival edge) emanating from a vertex arrived at by following a randomly chosen edge, are given by

$$\mathscr{H}_d(x,y;u;T_d,T_u)=\mathscr{H}_d(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u), \tag{10}$$

$$\mathscr{H}_r(x,y;u;T_d,T_u)=\mathscr{H}_r(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u), \tag{11}$$

$$\mathscr{H}_u(x, y; u; T_d, T_u) = \mathscr{H}_u(1 + (x-1)T_d, 1 + (y-1)T_d; 1 + (u-1)T_u). \tag{12}$$

2.3. Predicting the fate of a small outbreak

In general, percolation theory describes the behavior of connected groups of vertices in a random graph. We use the methods of percolation theory to predict the size of the infected cluster, that is, the number of vertices reached via disease transmission along the edges in the network. For a fixed network of contacts, there typically exists a threshold transmission rate below which only small, finite-sized outbreaks occur and above which large-scale epidemics (comparable to the size of the entire network) are possible.

We begin by deriving the value of the epidemic threshold and the expected size of small outbreaks below the threshold. These calculations assume that mildly contagious diseases spread in a tree-like fashion, causing only short transmission chains that do not loop back on themselves. We relax this assumption later, when we turn to diseases that lie above the epidemic threshold.

Let s denote the number of vertices contained in a small outbreak that begins at a randomly selected vertex. We now introduce symbols g and h which should not be confused with $\mathscr{G}$ and $\mathscr{H}$ (above), respectively. Let $g(w; T_d, T_u)$ be the generating function for the distribution of outbreak sizes:

$$g(w; T_d, T_u) = \sum_s P_s(T_d, T_u) w^s, \tag{13}$$

where $P_s(T_d, T_u)$ is probability that a single initial case sparks an outbreak of size s at the specified average transmissibilities. To solve for the average value of s, we first evaluate the size of an outbreak t that begins with a transmission event along a randomly chosen edge. If that edge is directed, then the set of vertices reached by occupied edges can be represented in graphical form as in the top row of Fig. 2. There are many possible outcomes: the disease does not spread along the edge, it spreads along the edge but no further, it spreads along the edge and then subsequently along another directed edge, it spreads along the edge and then subsequently along an undirected edge, it spreads along the original edge and then subsequently along two different directed edges emanating from the same vertex, etc. We will construct recursive equations to consider all possibilities.

We define a new generating function $h_d(w; T_d, T_u)$, which generates the probability distribution of t thus:

$$h_d(w; T_d, T_u) = \sum_t Q_t(T_d, T_u) w^t, \tag{14}$$

where $Q_t(T_d, T_u)$ is the probability that an outbreak beginning from a randomly chosen edge in the network will be size t. Fig. 2 illustrates that $h_d(w; T_d, T_u)$ satisfies a recursive condition of the form

$$h_d(w; T_d, T_u) = w\mathscr{H}_d(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u), \tag{15}$$

where $h_u(w)$ is the corresponding generating function for undirected edges, which itself satisfies a condition of the form

$$h_u(w; T_d, T_u) = w\mathscr{H}_u(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u) \tag{16}$$

as depicted in the bottom row of Fig. 2. The self-consistent solutions of Eqs. (15) and (16) give the distribution of t, given definitions (10) and (12) of $\mathscr{H}_d$ and $\mathscr{H}_u$. It follows that s, the size of an outbreak starting from a randomly chosen vertex, is distributed according to

$$g(w; T_d, T_u) = w\mathscr{G}(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u). \tag{17}$$

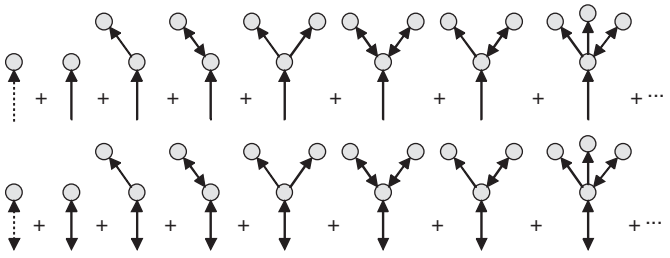

Fig. 2. Future transmission diagram. When disease is transmitted along a directed (top) or undirected (bottom) edge, we can consider all possible patterns of future transmission. Starting from a directed edge, for example, the disease may not spread along the edge, it may spread along the edge but no further, it may spread along the original edge and then subsequently along another directed edge, it may spread along the original edge and then subsequently along an undirected edge, it may spread along the original edge and then subsequently along two different directed edges emanating from the same vertex, etc. We construct recursive equations to consider all possible outcomes beginning from a single directed or undirected edge.

Consider now the average size of an outbreak starting from a random vertex $\langle s \rangle$, which is given by

$$\langle s \rangle = \sum_s s P_s(T_d, T_u) = g'(1; T_d, T_u), \tag{18}$$

where the prime denotes differentiation with respect to the first variable. In Appendix A.2, we derive the following expression for $\langle s \rangle$ in terms of the disease transmissibility and the pgf's for the degree distribution and excess degree distribution:

$$\langle s \rangle = 1 + \frac{T_d \mathscr{G}^{(0,1;0)}(1 - T_u(\mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)})) + T_u \mathscr{G}^{(0,0;1)}(1 - T_d(\mathscr{H}_d^{(0,1;0)} - \mathscr{H}_u^{(0,1;0)}))}{(1 - T_d \mathscr{H}_d^{(0,1;0)})(1 - T_u \mathscr{H}_u^{(0,0;1)}) - T_d T_u \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)}}, \tag{19}$$

where the arguments of all generating functions are set to (1,1;1). Thus, we can predict the expected size of an outbreak given the semi-directed degree distribution and transmissibilities T_d and T_u of the disease. If average transmissibilities along directed and undirected edges are equal ($T_d = T_u = T$), then the expected size of the outbreak is given by

$$\langle s \rangle = 1 + \frac{T \mathscr{G}^{(0,1;0)}(1 - T(\mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)})) + T \mathscr{G}^{(0,0;1)}(1 - T(\mathscr{H}_d^{(0,1;0)} - \mathscr{H}_u^{(0,1;0)}))}{(1 - T \mathscr{H}_d^{(0,1;0)})(1 - T \mathscr{H}_u^{(0,0;1)}) - T^2 \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)}}. \tag{20}$$

The expression for $\langle s \rangle$ diverges when the denominator in Eq. (19) is zero, and only predicts the expected size of the outbreak when the denominator is greater than zero. Thus the equation

$$(1 - T_d \mathscr{H}_d^{(0,1;0)})(1 - T_u \mathscr{H}_u^{(0,0;1)}) - T_d T_u \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)} = 0 \tag{21}$$

marks the phase transition at which the size of an outbreak first becomes extensive. Solving Eq. (21) for a given $0 \leqslant T_d \leqslant 1$, we derive the critical transmissibility T_{cu} at which a large-scale epidemic becomes possible:

$$T_{cu} = \frac{1 - T_d \mathscr{H}_d^{(0,1;0)}}{\mathscr{H}_u^{(0,0;1)} - T_d(\mathscr{H}_d^{(0,1;0)} \mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)})}. \tag{22}$$

Similarly, for some $0 \leqslant T_u \leqslant 1$, the critical value is defined by

$$T_{cd} = \frac{1 - T_u \mathscr{H}_u^{(0,0;1)}}{\mathscr{H}_d^{(0,1;0)} - T_u(\mathscr{H}_d^{(0,1;0)} \mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)})}. \tag{23}$$

Thus, there is a line defined by (22) and (23) of transmissibility values, below which we expect only small outbreaks of expected size $\langle s \rangle$ and above which an epidemic is possible. If average transmissibility is the same for directed and undirected edges, then there is a single critical transmissibility:

$$T_c = \frac{(\mathscr{H}_d^{(0,1;0)} + \mathscr{H}_u^{(0,0;1)}) \pm \sqrt{(\mathscr{H}_d^{(0,1;0)} + \mathscr{H}_u^{(0,0;1)})^2 - 4(\mathscr{H}_d^{(0,1;0)} \mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)})}}{2(\mathscr{H}_d^{(0,1;0)} \mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)} \mathscr{H}_u^{(0,1;0)})}, \tag{24}$$

whichever value is positive. We call $\{T_{cd}, T_{cu}\}$ (or T_c) the *epidemic threshold*. By differentiating Eqs. (6) and (8) and substituting the results into Eqs. (22) or (24) one can express the epidemic threshold in terms of the underlying structure of the contact network.

2.4. A simple example

We use these formulas to predict the spread of disease on a simple network in which all three degree distributions are Poisson with mean in-degree and out-degree of z_d and mean undirected degree of z_u. The pgf for the degree distribution is given by

$$\mathscr{G}(x, y; u) = \sum_{jkm} \left(\frac{z_d^j e^{-z_d}}{j!}\right)\left(\frac{z_d^k e^{-z_d}}{k!}\right)\left(\frac{z_u^m e^{-z_u}}{m!}\right) x^j y^k u^m = e^{z_d(x+y-2)+z_u(u-1)}. \tag{25}$$

The excess degree pgf's for this network are identical to the original degree distribution, that is, $\mathscr{H}_d(x, y; u) = \mathscr{H}_r(x, y; u) = \mathscr{H}_u(x, y; u) = \mathscr{G}(x, y; u)$. Therefore the expected size of an epidemic is defined by

$$\langle s \rangle = \frac{1}{1 - T_d z_d - T_u z_u}. \tag{26}$$

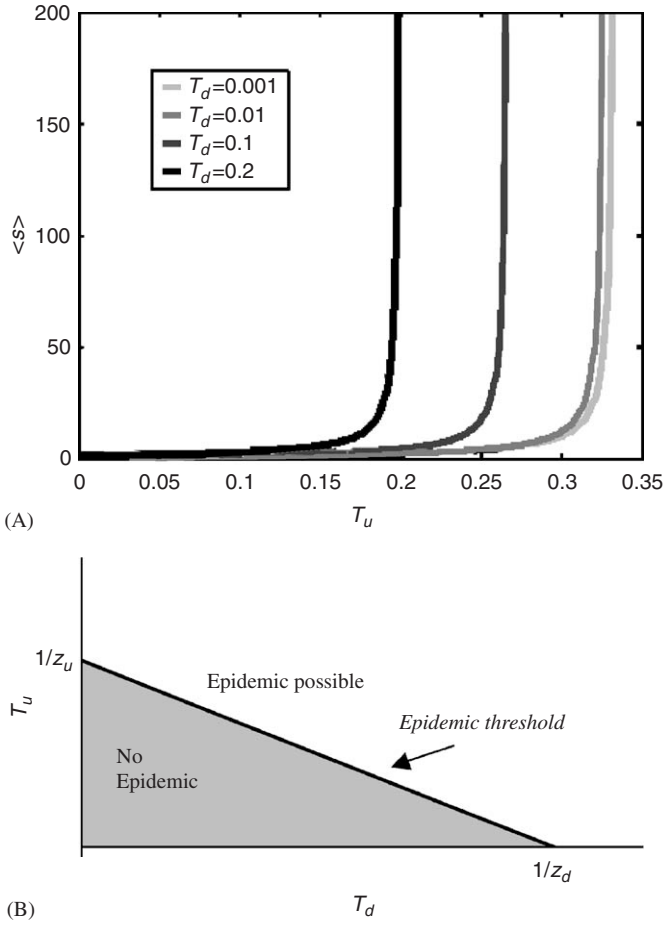

Fig. 3. Simple semi-directed network. (A) The expected size of a small outbreak as a function of T_d and T_u for a Poisson semi-directed network with Poisson parameters $z_d = 2$ and $z_u = 3$. (B) The epidemic threshold for a Poisson semi-directed network with Poisson parameters z_d and z_u.

We plot $\langle s \rangle$ for a Poisson semi-directed network in Fig. 3A. By setting the denominator equal to zero, we find an epidemic threshold line of

$$T_{cd} z_d + T_{cu} z_u = 1 \tag{27}$$

as depicted in Fig. 3B.

2.5. Probability and size of a large-scale epidemic

When the transmissibility of a disease is larger than the epidemic threshold, then Eq. (19) no longer indicates the size of the infected subpopulation. This is because transmission is so rampant that the chains of transmission are likely to loop back upon themselves, thus violating the assumption underlying the calculations depicted in Fig. 2.

When we are above the epidemic threshold, in the region in which epidemics can occur, we would like to know two quantities: the probability that a large-scale epidemic occurs and the fraction of individuals that are infected in that case. These quantities are equivalent to S_{in} and S_{out}—the fraction of vertices from which an extensive numbers of others can be reached by following occupied edges and the fraction of vertices contained in such an extensive interconnected group, respectively. In the language of percolation, these are the giant strongly connected component (GSCC) plus the giant in-component (GIN) and the GSCC plus the giant out-component (GOUT) defined by occupied edges. Fig. 4 illustrates the component structure of semi-directed networks. The relative size of the region shaded in vertical lines indicates the probability that any single infection will lead to a wide-spread epidemic, and the relative size of the region shaded in horizontal lines indicates the expected fraction of the population that will become infected during such an epidemic.

To calculate the typical size of a large-scale epidemic, we make use of the following argument. All vertices in the GSCC and GOUT are reachable from an extensive number of others (those in the GSCC and GIN), and all vertices that are not in these components are not reachable from an extensive number of others. We can calculate from how many vertices a randomly chosen vertex is reachable by following occupied edges *backwards* from that vertex and finding the resulting

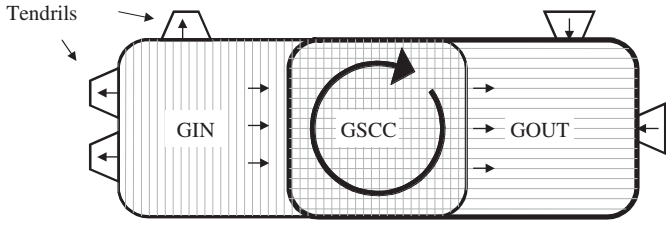

Fig. 4. Structure of a semi-directed network. The largest set of vertices for which you can move between any two by following edges in the correct direction is the *giant strongly connected component* (GSCC). The set of vertices not contained in the GSCC that can be reached by following edges in the correct direction from the GSCC is called the *giant out-component* (GOUT). The set of vertices not contained in the GSCC from which the GSCC can be reached by following edges in the correct direction is called the *giant in-component* (GIN). Vertices that are not in the GSCC, GIN, or GOUT but can either be reached from the GIN or can reach the GOUT are in the *tendrils* of the network.

component. This is precisely the reverse of the calculation we performed in the previous section, and allows us to derive the fraction of the graph contained in the largest occupied component, that is, the size of a large epidemic.

In Appendix A.3, we derive the following expression for the fraction of the population infected during a large-scale epidemic:

$$S_{\text{size}} = 1 - \sum_{jkm} p_{jkm}(1 + (a-1)T_d)^j(1 + (b-1)T_u)^m, \tag{28}$$

where a and b are solutions of

$$a = \frac{\sum_{jkm} k p_{jkm}(1 + (a-1)T_d)^j(1 + (b-1)T_u)^m}{\sum_{jkm} k p_{jkm}}, \tag{29}$$

$$b = \frac{\sum_{jkm} m p_{jkm}(1 + (a-1)T_d)^j(1 + (b-1)T_u)^{m-1}}{\sum_{jkm} m p_{jkm}}. \tag{30}$$

In most cases (28) is not solvable in closed form, but once we have the degree distribution (the p_{jkm}'s) and transmissibilities T_d and T_u it can be solved numerically by simple iteration, starting from appropriate initial values.

As discussed in Appendix A.3, one can similarly calculate the probability that an infection at a randomly chosen vertex will lead to a large-scale epidemic (S_{prob}). This quantity is equal to the size of the GSCC plus the GIN and is given by

$$S_{\text{prob}} = 1 - \sum_{jkm} p_{jkm}(1 + (\alpha-1)T_d)^k(1 + (\beta-1)T_u)^m, \tag{31}$$

where α and β are solutions of

$$\alpha = \frac{\sum_{jkm} j p_{jkm}(1 + (\alpha-1)T_d)^k(1 + (\beta-1)T_u)^m}{\sum_{jkm} j p_{jkm}}, \tag{32}$$

$$\beta = \frac{\sum_{jkm} m p_{jkm}(1 + (\alpha-1)T_d)^k(1 + (\beta-1)T_u)^{m-1}}{\sum_{jkm} m p_{jkm}}. \tag{33}$$

Note that α and β in Eqs. (31)–(33) are the probabilities that infection at a vertex at the end of a randomly selected directed and undirected edge (respectively) will not spark a large-scale epidemic. If average transmissibility is the same for directed and undirected edges, then simply substitute the single transmissibility value T for T_d and T_u in Eqs. (28), (29), (30), (31), (32), and (33).

These basic epidemiological quantities—the epidemic threshold and the fate of outbreaks on either side of the threshold—have been derived previously for completely directed (Newman, 2002) and completely undirected networks (Schwartz et al., 2002). We provide these formulae in Appendix A.4.

2.6. A simple example

Compartmental epidemiological models assume that the probability and expected size of an epidemic are always equal (Anderson and May, 1991; Hethcote, 2000). While this is true for undirected networks, these two values can be different in directed and semi-directed networks. We demonstrate this using three different networks: (N1) a completely undirected

Poisson network with mean degree z (where z is an even integer) that has generating function

$$\mathcal{G}_{N1}(x) = \sum_j \left(\frac{z^j e^{-z}}{j!}\right) x^j = e^{z(x-1)}, \tag{34}$$

(N2) a semi-directed network with a Poisson distribution of undirected edges of mean degree $z/2$, a Poisson in-degree distribution of mean $z/2$, and a regular out-degree distribution in which every vertex has an out-degree of exactly $z/2$ that has generating function

$$\begin{aligned}\mathcal{G}_{N2}(x,y;u) &= \sum_{jm} \left(\frac{(z/2)^j e^{-z/2}}{j!}\right)\left(\frac{(z/2)^m e^{-z/2}}{m!}\right) x^j y^{z/2} u^m \\ &= e^{(z/2)(x+u-2)} y^{z/2}\end{aligned} \tag{35}$$

and (N3) a fully directed network with a Poisson in-degree distribution of mean z, and a regular out-degree distribution in which every vertex has an out-degree of exactly z that has generating function

$$\mathcal{G}_{N3}(x,y) = \sum_j \left(\frac{z^j e^{-z}}{j!}\right) x^j y^z = e^{z(x-1)} y^z. \tag{36}$$

These three networks have the same total number of incoming and outgoing contacts since every undirected edge includes two incoming and two outgoing contacts while every directed edge includes a single incoming contact and a single outgoing contact. We use different in- and out-degree distributions to demonstrate the inequality of S_{prob} and S_{size}, because semi-directed and directed networks with identical in- and out-degree distributions have an equal-sized GIN and GOUT and therefore equal values of S_{prob} and S_{size}.

All three networks share the same epidemic threshold of $T_c = 1/z$. Above the threshold, the probabilities and expected sizes of epidemics in these networks are predicted by the following equations:

$$S_{N1} = S_{N2_{size}} = S_{N3_{size}} = 1 - e^{z(u-1)T} \quad \text{where } u = e^{z(u-1)T}, \tag{37}$$

$$S_{N2_{prob}} = 1 - (1 + (\alpha - 1)T)^{z/2} e^{z/2(\alpha-1)T} \quad \text{where } \alpha = (1 + (\alpha - 1)T)^{z/2} e^{z/2(\alpha-1)T}, \tag{38}$$

$$S_{N3_{prob}} = 1 - (1 + (w - 1)T)^z, \quad \text{where } w = (1 + (w - 1)T)^z. \tag{39}$$

Fig. 5 illustrates these predictions for two sets of networks ($z = 4$ and 8). For each set of networks, all three share the same expected size of an epidemic. The probability of an epidemic is identical to the expected size of an epidemic in the undirected network, much larger than the expected size in the completely directed network, and at an intermediate value in the semi-directed network. Our particular choice of in- and out-degree distributions yields networks with GIN larger than GOUT. If we reverse these two distributions, then GOUT would be larger than GIN, and therefore, the expected size of the epidemic would be larger than the probability of an epidemic.

2.7. Initial conditions

We can refine our predictions if we know something about the behavior of *patient zero*—the first case of disease in a population. Suppose, for instance, that we know patient zero has out-degree k and undirected-degree m. The probability that he or she will spark a large-scale epidemic is just the probability that transmission of the disease along at least one of the edges emanating from patient zero will lead to an epidemic. For any one of its k out edges and m undirected edges, the probability that the disease is not transmitted along the edge is $1 - T_d$ and $1 - T_u$, respectively. As defined in Eqs. (32) and (33), α and β are the probabilities that an outbreak traveling along a given directed or undirected edge will spread to only a local component of the population. Thus the probability that disease is transmitted along one of the $k + m$ edges but does not proceed from there into a full-blown epidemic is $T_d\alpha$ for a directed edge or $T_u\beta$ for an undirected edge, and the overall probability that patient zero will spark an epidemic is given by

$$\varepsilon_{km} = 1 - (1 - T_d + T_d\alpha)^k (1 - T_u + T_u\beta)^m. \tag{40}$$

The probability that an outbreak of size N will lead to a large-scale epidemic is $1 - \prod_{i=1}^{N}(1 - \varepsilon_{k_i m_i})$, where k_i is the out-degree and m_i is the undirected-degree of individual i. This is just one minus the probability that none of the N infected individuals sparks an epidemic. If we know the number of current cases but not their contact patterns, then our best estimate for the probability of an epidemic is calculated similarly, with each of the $(1 - \varepsilon_{k_i m_i})$'s replaced with the probability that a typical infected individual does not spark an epidemic. Such an individual was infected either along a directed edge with a priori probability $z_d/(z_d + z_u)$ or along an undirected edge with a priori probability $z_u/(z_d + z_u)$. The number of

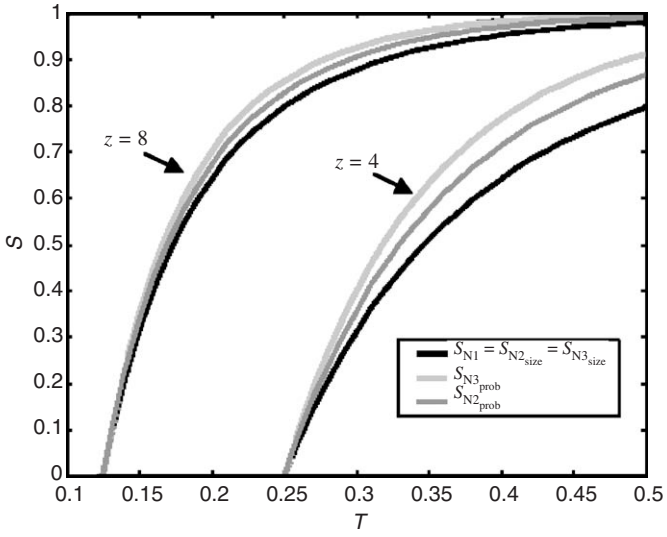

Fig. 5. Epidemiological predictions for undirected, directed and semi-directed networks. The probability of an epidemic and expected fraction of the population infected during an epidemic for three classes networks: (N1) a completely undirected Poisson network with mean degree z; (N2) a semi-directed network with Poisson undirected and in-degree distributions of mean degree $z/2$, and with every vertex having an out-degree of exactly $z/2$; and (N3) a completely directed network with a Poisson in-degree distribution of mean degree z, and with every vertex having an out-degree of exactly z. For each of these networks, we plot the predicted probability and size of an epidemic (S) as a function of the average transmissibility of the disease (T). S_{N1} is both the expected magnitude and probability of an epidemic in network N1 and $S_{N2_{size}}$ and $S_{N2_{prob}}$ ($S_{N3_{size}}$ and $S_{N3_{prob}}$) are the expected magnitude and probability of an epidemic in N2 (N3). The left and right three lines correspond to networks with $z = 8$ and 4, respectively.

edges through which the individual can start an epidemic is given by the excess degree generating functions $\mathscr{H}_d$ and $\mathscr{H}_u$, and the probability that one of those edges will not give rise to an epidemic is $1 - T_d + T_d\alpha$ for a directed edge and $1 - T_u + T_u\beta$ for an undirected edge. Thus the probability that a typical infected individual does not start an epidemic is given by

$$\frac{z_d \mathscr{H}_d(1,(1-T_d+T_d\alpha);(1-T_u+T_u\beta)) + z_u \mathscr{H}_u(1,(1-T_d+T_d\alpha);(1-T_u+T_u\beta))}{z_d+z_u}, \tag{41}$$

and the probability that an outbreak of size N sparks an epidemic is given by

$$1 - \left(\frac{z_d \mathscr{H}_d(1,(1-T_d+T_d\alpha);(1-T_u+T_u\beta)) + z_u \mathscr{H}_u(1,(1-T_d+T_d\alpha);(1-T_u+T_u\beta))}{z_d+z_u}\right)^N \tag{42}$$

where α and β are as described by Eqs. (32) and (33). Simplifying slightly, we can rewrite Eq. (42) as

$$1 - \left(\frac{\sum_{jkm} j p_{jkm}(1-T_d+T_d\alpha)^k(1-T_u+T_u\beta)^m + \sum_{jkm} m p_{jkm}(1-T_d+T_d\alpha)^k(1-T_u+T_u\beta)^{m-1}}{\sum_{jkm}(j+m)p_{jkm}}\right)^N. \tag{43}$$

Appendix A.4 provides the analogous equations for completely directed and completely undirected networks.

2.8. Individual risk and intervention

The likelihood that an individual of in-degree j and undirected-degree m will be infected during an epidemic is equal to one minus the probability that none of his or her $j + m$ contacts will transmit the disease to him or her. The probability that a contact does not transmit the disease is equal to the probability that the contact was infected, but did not transmit the disease, $1 - T_d$ for a contact along a directed edge and $1 - T_u$ for a contact along an undirected edge, plus the probability that the contact was not infected in the first place, $T_d a$ for a contact along a directed edge or $T_u b$ for a contact along an undirected edge, where a and b are as defined by Eqs. (29) and (30). Thus, a randomly chosen vertex of in-degree j and undirected-degree m will become infected with probability

$$v_{jm} = 1 - (1 - T_d + T_d a)^j (1 - T_u + T_u b)^m. \tag{44}$$

When a single individual of degrees j, k, m lowers the likelihood of transmission to or from himself or herself (by wearing a face mask in the case of an air-borne disease, for example) from T_d and T_u to $\phi_d T_d$ and $\phi_u T_u$ ($0 \leqslant \phi_d, \phi_u \leqslant 1$), then the expressions for the likelihood of causing an epidemic and becoming infected during an epidemic become

$$\varepsilon_{km}^{\phi} = 1 - (1 - \phi_d T_d + \phi_d T_d \alpha)^k (1 - \phi_u T_u + \phi_u T_u \beta)^m, \tag{45}$$

$$v_{jm}^{\phi} = 1 - (1 - \phi_d T_d + \phi_d T_d a)^j (1 - \phi_u T_u + \phi_u T_u b)^m, \tag{46}$$

where a, b, α, and β are as in Eqs. (29), (30), (32), and (33).

Note that these two quantities are different for some semi-directed networks, whereas they are always identical for undirected networks (Appendix A.4).

3. A case study in hospital-based transmission of respiratory disease

3.1. The contact networks

We have previously developed a method to simulate urban contact networks based on demographic data for the city of Vancouver, British Columbia (Statistics Canada, 2001; BC Stats, 2002; Centre for Health Sevices and Policy Research, 2002; Vancouver School Board, 2002; BC Stats, 2003; Meyers et al., 2005). Using the degree distribution from a contact network model containing 10,000 households (~25,000 individuals), we predict the fate of an outbreak for a spectrum of respiratory-borne diseases for which hospitalization is likely. As reported in (Meyers et al., 2005), the undirected-degree distribution is roughly exponential. The in-degree and out-degree distributions are solely determined by the flow of infected people into health care facilities. In this model, we make the simple assumption that each non-HCW member of the population has three directed edges pointing to randomly chosen HCWs in his or her local hospital. Thus a typical individual has out-degree of three and in-degree of zero; and a typical HCW has out-degree of zero and in-degree ranging from 409 to 530. Because the mode of transmission (respiratory-borne) is the same for directed and undirected edges in this network, we assume that there is a single average transmissibility across the entire network, that is, $T_d = T_u = T$.

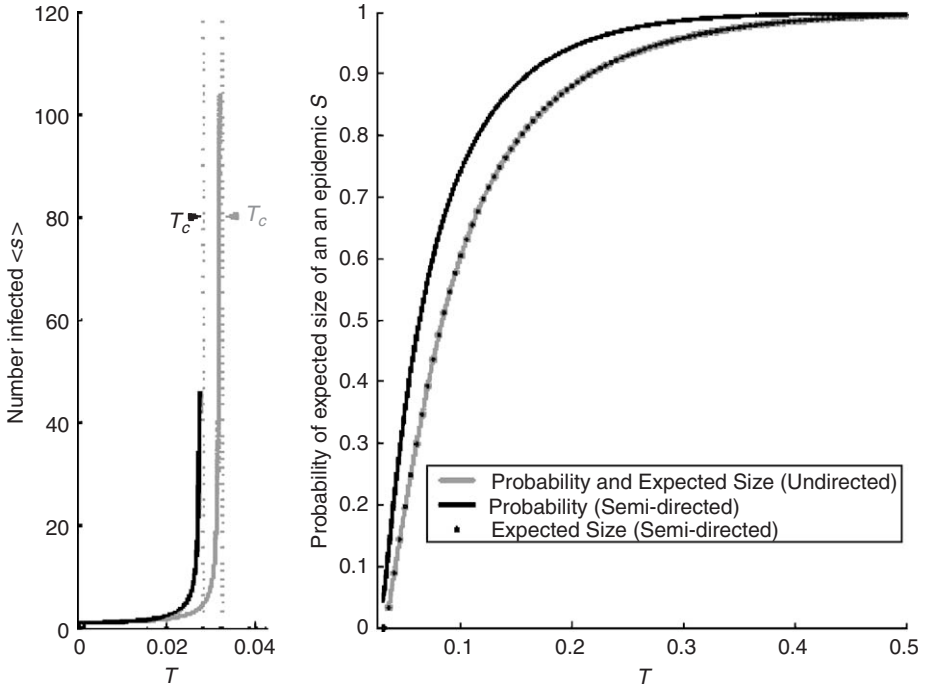

Fig. 6. Epidemiological predictions on undirected and semi-directed contact networks. This graph shows the expected size of small outbreaks below the epidemic threshold (left), and the probability and expected size of a large-scale epidemic above the epidemic threshold (rate) for diseases with various transmission rates (T) spreading through an urban contact network. The predictions for the semi-directed and undirected networks are shown in black and gray, respectively.

(This assumption would not be appropriate for a disease in which directed and undirected edges represented different modes of transmission.)

Using the formulae derived above, we calculated the epidemic threshold for this particular contact network, the expected size of an outbreak for diseases below the epidemic threshold, and the probability and expected size of an epidemic for diseases above the epidemic threshold (Fig. 6). The inclusion of one-directional disease transmission from the general public into health care settings significantly increases the vulnerability of a population. The epidemic threshold is lowered from $T_c = 0.0322$ to 0.0278. The lower the transmissibility of the disease, the more pronounced the impact of hospital-based transmission. For diseases close to the epidemic threshold, the probability of an epidemic in the more realistic semi-directed network is more than double that of the simpler undirected network. Note that if an epidemic does ensue, the expected size of an epidemic is almost identical for the two contact networks.

We can also predict the role of HCWs in the spread of disease and the impact of intervention. There are two basic categories of intervention (Meyers et al., 2005; Pourbohloul et al., 2005). *Contact reducing* interventions modify the basic patterns of interaction. Within hospitals, for example, suspected cases are isolated in negative pressure rooms and the number of caregivers attending to such patients is limited. For the population at large, public health officials may implement quarantines and travel restrictions. Such interventions can be modeled by removing appropriate edges from the contact network. Vaccination prior to an outbreak, which entails removing a vertex and all of its edges from the contact network, is the extreme form of such interventions. *Transmission reducing* interventions like the use of facemasks, surgical gowns, and hand washing lower the probability of infecting existing contacts.

During an outbreak of a new infectious disease, the patient burden to hospitals may be so severe that health care officials cannot reasonably lower the number of contacts between HCWs and patients. Instead, as with SARS, they often implement strict hygienic precautions that lower transmissibility (Le et al., 2003; McDonald et al., 2004). Fig. 7 illustrates the impact of various levels of transmission reducing interventions within hospitals. Here we assume that the average transmission rate along directed edges only (T_d) is reduced. This models hygienic precautions taken by HCWs while treating suspected cases of the disease. If a HCW becomes infected, the threat remains high because of a large number of

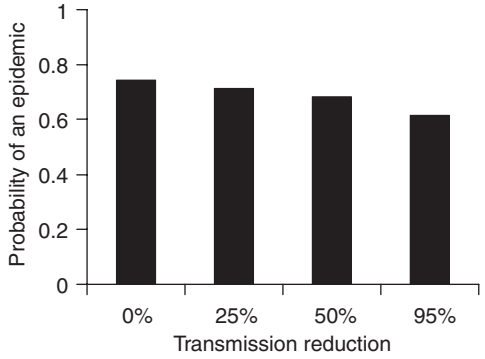

Fig. 7. Hospital-based intervention. The probability of a large-scale epidemic decreases as the HCWs use increasingly strict hygienic precautions for a disease originally above the epidemic threshold, $T_d = T_u = 0.1$. The x-axis gives the percent reduction in transmissibility along directed edges pointing from members of the general public to HCWs.

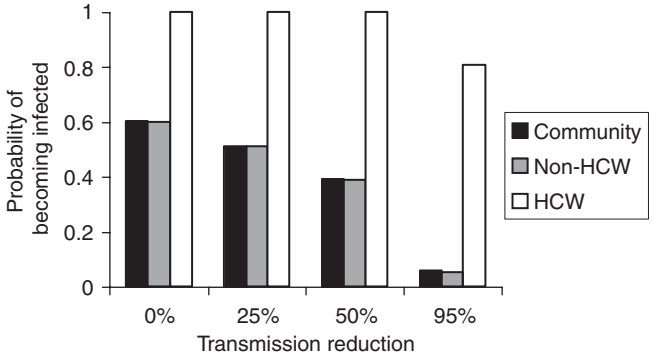

Fig. 8. Individual precautions. An individual can lower the probability that he or she will become infected during an epidemic by taking measures that limit transmission. The x-axis gives the percent reduction in transmissibility between the individual and all of his or her directed and undirected contacts for a disease originally above the epidemic threshold, $T_d = T_u = 0.1$. The average likelihood of infection across the entire population is shown in black bars. The benefit of intervention is much greater for members of the general public (gray bars) than for HCWs (white bars).

uncontrolled undirected contacts with other HCWs and patients who are hospitalized for other conditions. These measures would therefore be more effective if they were extended to all HCW–patient interactions. In sum, the use of transmission reducing interventions by HCWs treating suspected cases will protect the population only when they block transmission to hospital personnel entirely.

In the absence of organized intervention, individuals may choose to take precautions. Before much was known about SARS, HCWs made individual choices about prevention, and later in the outbreak, some members of the general public voluntarily wore facemasks (Tang and Wong, 2004). In Fig. 8, we show the personal impact of such precautions. On average, taking drastic transmission reducing measures can significantly lower the probability of becoming infected during an epidemic. Yet HCWs are not nearly as protected by such individual measures as are members of the general public. This stems from the sheer numbers of potential contacts between HCWs and infected patients.

4. Discussion

We have derived a number of important epidemiological quantities for semi-directed contact networks in which the average transmissibility can be different for directed and undirected contacts. When there are, in fact, two different transmission rates, the epidemic threshold becomes a line dividing the space of transmission rates into a region in which there are only small outbreaks that die out before reaching a sizable fraction of the population and another region in which an epidemic is possible (Fig. 3).

Above the epidemic threshold, semi-directed networks are more complicated than undirected networks. When the in-degree and out-degree distributions differ, then so do the probability of an epidemic and the expected incidence should one occur. We have illustrated the differences between undirected, semi-directed, and directed networks using three simple networks that share the same total number of contacts. The gap between the probability and expected size of an epidemic is non-existent for the undirected networks, quite large for the directed network, and somewhere in between for the semi-directed network (Fig. 5). In addition to these fundamental epidemiological quantities, we have also calculated the probability of an epidemic as a function of the degree of the first case and the impact of control measures on the complying individual and the population as a whole.

We have applied these methods to study the pivotal role of hospitals in the spread of air-borne diseases through communities. Worldwide outbreaks of SARS between November 2002 and May 2003 increased public awareness about the devastating human, economic and psychological impact of emerging infectious diseases. SARS probably emerged in Southern China from an animal reservoir and was transmitted primarily through respiratory droplets and secondarily through aerosolized gastrointestinal secretions (Donnelly et al., 2003). From the beginning, SARS exhibited distinctive epidemiological patterns. During its initial four months of spread in China, 32% of confirmed cases were HCWs and 39% were food handlers (hence the hypothesis that cooking wild animals was the primary route of SARS transmission into human populations), yet there were no cases among schoolchildren or housewives (Xu et al., 2004).

As SARS spread out from China, the fate of outbreaks was tightly linked to containment efforts within hospitals (Le et al., 2003; McDonald et al., 2004). For example, the first cases of SARS in Vancouver and Toronto were infected almost simultaneously while staying in Hotel M in Hong Kong. Whereas the Toronto case sparked a sizeable outbreak that involved extensive hospital-based transmission, no secondary cases occurred from the initial Vancouver case. The successful containment in Vancouver may have stemmed from rigorous hospital precautions. In particular, the Vancouver emergency room at which the first case sought treatment had recently participated in an infection control audit that emphasized the importance of barrier precautions for all acute onset respiratory infections (World Health Organization, 2003, 2004; Skowronski et al., 2005). In contrast, patient zero in Toronto died at home as an undiagnosed case of SARS after infecting several relatives. The first case to arrive in a Toronto hospital (on March 7, 2003) was a second-generation, locally acquired case. He was treated with nebulized salbutamol in the emergency room, where he remained for 18 h without special precautions. After 21 h, he was placed in air-borne isolation in the ICU for possible tuberculosis, and droplet and contact precautions were not applied until his fourth day in the hospital. By the time he died on March 13, he had infected several HCWs and thereby exacerbated the Toronto outbreak (Poutanen et al., 2003; Varia et al., 2003).

Given the importance of hospitals to the transmission and control of diseases like SARS, we have developed a mathematical framework that explicitly models the flow of patients into hospitals during outbreaks. In particular, we demonstrate that a semi-directed contact network can capture a conditional contact between a layperson and a HCW that only occurs if the layperson becomes infected and goes to the hospital. We can rapidly predict the spread of disease through such a network using an extension of the methods developed in (Newman, 2002) and (Meyers et al., 2005).

When we add interactions between HCWs and infected patients to our model, the predicted epidemic threshold—the critical transmission rate above which outbreaks may evolve into full-blown epidemics—decreases and the risk of infection for HCWs dramatically increases. Furthermore, interventions targeted at reducing the likelihood of transmission from patients and HCWs may significantly lower the likelihood of an epidemic. Thus, models that ignore hospital-based

transmission may underestimate both the threat of an epidemic and the impact of control measures targeted at protecting HCWs from the onslaught of patients they may face during an outbreak.

The extension of contact network epidemiology to semi-directed graphs will allow us to build and rapidly analyse more realistic models of infectious disease transmission when there are asymmetries in disease causing contacts. As our hospital example demonstrates, such models may provide important new insights into epidemiological patterns and public health strategy.

Acknowledgments

The authors would like to thank Robert Brunham and Danuta Skowronski of the University of British Columbia Centre for Disease Control for fruitful discussions. This work was supported in part by grants from the Canadian Institutes of Health Research (CIHR) to the Canadian SARS Investigators that includes L.A.M and B.P. (FRN: 67803), from the National Science Foundation (DEB-0303636) to L.A.M., and from the James S. McDonnell Foundation and the National Science Foundation (DMS-0234188 and DMS-0405348) to M.E.J.N. The Santa Fe Institute and CIHR supported the working visits of B.P., L.A.M., and M.E.J.N.

Appendix A

A.1. Probability generating functions for the number of occupied (infected) edges

Following Newman (2002), we write the generating function for the number of occupied edges of a vertex in the form

$$\begin{aligned}\mathscr{G}(x,y;u;T_d,T_u) &= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\left[\sum_{j=a}^{\infty}\sum_{k=b}^{\infty}\sum_{m=c}^{\infty}p_{jkm}\binom{j}{a}T_d^a(1-T_d)^{j-a}\binom{k}{b}T_d^b(1-T_d)^{k-b}\binom{m}{c}T_u^c(1-T_u)^{m-c}x^a y^b u^c\right]\\ &= \sum_{jkm}p_{jkm}\left[\sum_{a=0}^{j}\binom{j}{a}(T_d x)^a(1-T_d)^{j-a}\sum_{b=0}^{k}\binom{k}{b}(T_d y)^b(1-T_d)^{k-b}\sum_{c=0}^{m}\binom{m}{c}(T_u u)^c(1-T_u)^{m-c}\right].\end{aligned} \tag{47}$$

Applying the binomial formula ($(a+b)^n=\sum_{k=0}^{n}a^{n-k}b^k$), this generating function simplifies to

$$\begin{aligned}\mathscr{G}(x,y;u;T_d,T_u) &= \sum_{jkm}p_{jkm}(1-T_d+xT_d)^j(1-T_d+yT_d)^k(1-T_u+uT_u)^m\\ &= \mathscr{G}(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u).\end{aligned} \tag{48}$$

We similarly derive the probability generating function for the number of occupied edges (excluding the arrival edge) emanating from a vertex arrived at by following a randomly chosen edge:

$$\mathscr{H}_d(x,y;u;T_d,T_u)=\mathscr{H}_d(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u), \tag{49}$$

$$\mathscr{H}_r(x,y;u;T_d,T_u)=\mathscr{H}_r(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u), \tag{50}$$

$$\mathscr{H}_u(x,y;u;T_d,T_u)=\mathscr{H}_u(1+(x-1)T_d,1+(y-1)T_d;1+(u-1)T_u). \tag{51}$$

Eq. (48) implies that

$$\mathscr{G}(x,y;u;1,1)=\mathscr{G}(x,y;u), \tag{52}$$

$$\mathscr{G}(1,1;1;T_d,T_u)=\mathscr{G}(1,1;1), \tag{53}$$

$$\mathscr{G}^{(1,0;0)}(1,1;1;T_d,T_u)=T_d\mathscr{G}^{(1,0;0)}(1,1;1), \tag{54}$$

$$\mathscr{G}^{(0,1;0)}(1,1;1;T_d,T_u)=T_d\mathscr{G}^{(0,1;0)}(1,1;1), \tag{55}$$

$$\mathscr{G}^{(0,0;1)}(1,1;1;T_d,T_u)=T_u\mathscr{G}^{(0,0;1)}(1,1;1) \tag{56}$$

and similarly for $\mathscr{H}_d$, $\mathscr{H}_r$, and $\mathscr{H}_u$.

A.2. Derivation of the size of a small outbreak

As explained in the text, the generating functions for the size of an outbreak beginning with a randomly chosen *edge* are given by

$$h_d(w; T_d, T_u) = w\mathscr{H}_d(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u), \tag{57}$$

$$h_u(w; T_d, T_u) = w\mathscr{H}_u(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u) \tag{58}$$

for directed and undirected starting edges, respectively. Furthermore, the generating function for the size of an outbreak starting from a randomly chosen *vertex* is given by

$$g(w; T_d, T_u) = w\mathscr{G}(1, h_d(w; T_d, T_u); h_u(w; T_d, T_u); T_d, T_u). \tag{59}$$

Here we derive the expected size of such an outbreak

$$\langle s\rangle = \sum_s sP_s(T_d, T_u) = g'(1; T_d, T_u), \tag{60}$$

where the prime denotes differentiation with respect to the first variable. Differentiating equations (15), (16), and (17), we find

$$g'(1; T_d, T_u) = 1 + \mathscr{G}^{(0,1;0;0)}(1, 1; 1; T_d, T_u)h'_d(1; T_d, T_u) + \mathscr{G}^{(0,0;1;0)}(1, 1; 1; T_d, T_u)h'_u(1; T_d, T_u), \tag{61}$$

$$h'_d(1; T_d, T_u) = 1 + \mathscr{H}_d^{(0,1;0;0)}(1, 1; 1; T_d, T_u)h'_d(1; T_d, T_u) + \mathscr{H}_d^{(0,0;1;0)}(1, 1; 1; T_d, T_u)h'_u(1; T_d, T_u), \tag{62}$$

$$h'_u(1; T_d, T_u) = 1 + \mathscr{H}_u^{(0,1;0;0)}(1, 1; 1; T_d, T_u)h'_d(1; T_d, T_u) + \mathscr{H}_u^{(0,0;1;0)}(1, 1; 1; T_d, T_u)h'_u(1; T_d, T_u), \tag{63}$$

where we have made use of the fact that $h_d(1; T_d, T_u) = h_u(1; T_d, T_u) = 1$. Solving Eqs. (62) and (63) simultaneously we find

$$h'_d(1; T_d, T_u) = \frac{1 - \mathscr{H}_u^{(0,0;1;0)} + \mathscr{H}_d^{(0,0;1;0)}}{(1 - \mathscr{H}_d^{(0,1;0;0)})(1 - \mathscr{H}_u^{(0,0;1;0)}) - \mathscr{H}_d^{(0,0;1;0)}\mathscr{H}_u^{(0,1;0;0)}}, \tag{64}$$

$$h'_u(1; T_d, T_u) = \frac{1 - \mathscr{H}_d^{(0,1;0;0)} + \mathscr{H}_u^{(0,1;0;0)}}{(1 - \mathscr{H}_d^{(0,1;0;0)})(1 - \mathscr{H}_u^{(0,0;1;0)}) - \mathscr{H}_d^{(0,0;1;0)}\mathscr{H}_u^{(0,1;0;0)}}, \tag{65}$$

where the arguments of all generating functions are set to (1,1;1;T_d,T_u). Substituting these expressions into Eq. (61), we calculate the expected size of an outbreak beginning at a random vertex:

$$\langle s\rangle = 1 + \frac{\mathscr{G}^{(0,1;0;0)}(1 - \mathscr{H}_u^{(0,0;1;0)} + \mathscr{H}_d^{(0,0;1;0)}) + \mathscr{G}^{(0,0;1;0)}(1 - \mathscr{H}_d^{(0,1;0;0)} + \mathscr{H}_u^{(0,1;0;0)})}{(1 - \mathscr{H}_d^{(0,1;0;0)})(1 - \mathscr{H}_u^{(0,0;1;0)}) - \mathscr{H}_d^{(0,0;1;0)}\mathscr{H}_u^{(0,1;0;0)}}. \tag{66}$$

Eqs. (52)–(56) allow us to separate transmissibilities T_d and T_u from the semi-directed degree distributions as follows:

$$\langle s\rangle = 1 + \frac{T_d\mathscr{G}^{(0,1;0)}(1 - T_u(\mathscr{H}_u^{(0,0;1)} - \mathscr{H}_d^{(0,0;1)})) + T_u\mathscr{G}^{(0,0;1)}(1 - T_d(\mathscr{H}_d^{(0,1;0)} - \mathscr{H}_u^{(0,1;0)}))}{(1 - T_d\mathscr{H}_d^{(0,1;0)})(1 - T_u\mathscr{H}_u^{(0,0;1)}) - T_dT_u\mathscr{H}_d^{(0,0;1)}\mathscr{H}_u^{(0,1;0)}}, \tag{67}$$

where the arguments of all generating functions are now set to (1,1;1).

A.3. Derivation of the size and probability of a large epidemic

Suppose we start at a randomly chosen edge and move backwards along directed edges and along undirected edges, traversing each directed and undirected edge with probability T_d or T_u, respectively. Then the distribution of the sizes of the resulting components is generated by $h_r(w; T_d, T_u)$ (if starting from a random directed edge) and $h_{ur}(w; T_d, T_u)$ (if starting from a random undirected edge) which satisfy

$$h_r(w; T_d, T_u) = w\mathscr{H}_r(h_r(w; T_d, T_u), 1; h_{ur}(w; T_d, T_u); T_d, T_u) \tag{68}$$

and

$$h_{ur}(w; T_d, T_u) = w\mathscr{H}_u(h_r(w; T_d, T_u), 1; h_{ur}(w; T_d, T_u); T_d, T_u). \tag{69}$$

It follows that the distribution of components from which a randomly chosen vertex (rather than edge) can be reached is generated by

$$g_r(w;T_d,T_u)=w\mathcal{G}(h_r(w;T_d,T_u),1;h_{ur}(w;T_d,T_u);T_d,T_u). \tag{70}$$

The fraction of the graph filled by vertices for which the corresponding component is finite in size is then given by $g_r(1;T_d,\ T_u)$ and hence $S_{\text{out}}=1-g_r(1;T_d,T_u)$ giving

$$S_{out}=1-\mathcal{G}(a,1;b;T_d,T_u), \tag{71}$$

where $a\equiv h_r(1;T_d,T_u)$ and $b\equiv h_{ur}(1;T_d,T_u)$ are solutions of

$$a=\mathcal{H}_r(a,1;b;T_d,T_u),\quad b=\mathcal{H}_u(a,1;b;T_d,T_u). \tag{72}$$

Translating into epidemiological terms, we can predict the size of a large-scale epidemic from the degree distribution and transmissibility with

$$S_{size}=S_{out}=1-\sum_{jkm}p_{jkm}(1+(a-1)T_d)^j(1+(b-1)T_u)^m, \tag{73}$$

where a and b are solutions of

$$a=\frac{\sum_{jkm}kp_{jkm}(1+(a-1)T_d)^j(1+(b-1)T_u)^m}{\sum_{jkm}kp_{jkm}}, \tag{74}$$

$$b=\frac{\sum_{jkm}mp_{jkm}(1+(a-1)T_d)^j(1+(b-1)T_u)^{m-1}}{\sum_{jkm}mp_{jkm}}. \tag{75}$$

Similarly, one can calculate S_{in}, the size of the GSCC plus the GIN, which is the fraction of vertices from which an extensive number of others can be reached. In epidemiology, this is S_{prob}, the probability that a single randomly placed infection will spark a large-scale epidemic. By analogy with Eqs. (71) and (72), S_{in} is given by

$$S_{in}=1-\mathcal{G}(1,\alpha;\beta;T_d,T_u), \tag{76}$$

where

$$\alpha=\mathcal{H}_d(1,\alpha;\beta;T_d,T_u),\quad \beta=\mathcal{H}_u(1,\alpha;\beta;T_d,T_u). \tag{77}$$

In terms of the degree distribution and transmissibility of disease, Eqs. (76) and (77) become

$$S_{prob}=S_{in}=1-\sum_{jkm}p_{jkm}(1+(\alpha-1)T_d)^k(1+(\beta-1)T_u)^m, \tag{78}$$

where α and β are solutions of

$$\alpha=\frac{\sum_{jkm}jp_{jkm}(1+(\alpha-1)T_d)^k(1+(\beta-1)T_u)^m}{\sum_{jkm}jp_{jkm}}, \tag{79}$$

$$\beta=\frac{\sum_{jkm}mp_{jkm}(1+(\alpha-1)T_d)^k(1+(\beta-1)T_u)^{m-1}}{\sum_{jkm}mp_{jkm}}. \tag{80}$$

A.4. Epidemic quantities for completely directed and completely undirected graphs

A.4.1. Basic quantities

The epidemic threshold, expected size of a small outbreak, probability of a large scale epidemic and the expected size of such an epidemic have been derived previously for both undirected networks (Newman, 2002) and completely directed networks (Schwartz et al., 2002). For directed networks, the expected size of a small outbreak is given by

$$\langle s\rangle=1+\frac{T\mathcal{G}^{(0,1)}(1,1)}{1-T\mathcal{H}_d^{(0,1)}(1,1)} \tag{81}$$

when the denominator on the right-hand side is greater than zero. This yields an epidemic threshold of

$$T_c = \frac{1}{\mathscr{H}_d^{(0,1)}(1,1)}. \tag{82}$$

For transmissibility above T_c, the expected size of an epidemic is given by

$$S_{size}^d = 1 - \sum_{jk} p_{jk}(1 + (v-1)T)^j, \tag{83}$$

where v is the solution to the equation

$$v = \frac{\sum_{jk} k p_{jk}(1 + (v-1)T)^j}{\sum_{jk} k p_{jk}} \tag{84}$$

and the probability that such an epidemic will arise in the first place is given by

$$S_{prob}^d = 1 - \sum_{jk} p_{jk}(1 + (w-1)T)^k, \tag{85}$$

where w is the solution to the equation

$$w = \frac{\sum_{jk} j p_{jk}(1 + (w-1)T)^k}{\sum_{jk} j p_{jk}}. \tag{86}$$

For undirected networks, the expected size of a small outbreak is given by

$$\langle s \rangle = 1 + \frac{T\mathscr{G}'(1)}{1 - T\mathscr{H}'(1)} \tag{87}$$

with an epidemic threshold of

$$T_c = \frac{1}{\mathscr{H}'(1)}. \tag{88}$$

For $T > T_c$, the probability and expected size of an epidemic in an undirected network are identical and given by

$$S^u = S_{size}^u = S_{prob}^u = 1 - \sum_m p_m(1 + (u-1)T)^m, \tag{89}$$

where u is the solution to the equation

$$u = \frac{\sum_m m p_m(1 + (u-1)T)^{m-1}}{\sum_m m p_m}. \tag{90}$$

A.4.2. Initial conditions and individual risk

First consider a completely directed network. If patient zero has out-degree k, then the probability that he or she will spark a large-scale epidemic is given by

$$\varepsilon_k = 1 - (1 - T + Tw)^k, \tag{91}$$

where w is the solution to (86). If he or she complies with an intervention that lowers the probability of transmission to others by a factor $\phi (0 \leqslant \phi \leqslant 1)$, then the probability of sparking an epidemic is reduces to

$$\varepsilon_k^\phi = 1 - (1 - \phi T + \phi T w)^k. \tag{92}$$

If there is an initial outbreak of N cases (of unknown degree), then the probability of a large-scale epidemic is given by

$$1 - \left(\frac{\sum_{jk} j p_{jk}(1 - T + Tw)^k}{\sum_{jk} j p_{jk}} \right)^N, \tag{93}$$

where w is the solution to (86).

During a large-scale epidemic, the probability that a randomly chosen vertex of in-degree j will become infected is given by

$$v_j = 1 - (1 - T + Tv)^j, \tag{94}$$

where v is the solution to (84). If he or she complies with an intervention that lowers the likelihood of transmission from infected contacts by a factor $\phi(0 \leqslant \phi \leqslant 1)$, then the probability of infection is reduced to

$$v_j = 1 - (1 - \phi T + \phi T v)^j. \tag{95}$$

We next give the analogous formulae for undirected networks. These were originally derived in Meyers et al. (2005). If patient zero has degree k, then the probability that he or she will spark a large-scale epidemic is given by

$$\varepsilon_k = 1 - (1 - T + Tu)^k, \tag{96}$$

where u is the solution to (90). If he or she complies with an intervention that lowers the probability of transmission to others by a factor $\phi(0 \leqslant \phi \leqslant 1)$, then the probability of sparking an epidemic is reduces to

$$\varepsilon_k^{\phi} = 1 - (1 - \phi T + \phi T u)^k. \tag{97}$$

If there is an initial outbreak of N cases (of unknown degree), then the probability of a large-scale epidemic is given by

$$1 - \left(\frac{\sum_k k p_k (1 - T + Tu)^{k-1}}{\sum_k k p_k} \right)^N, \tag{98}$$

where u is the solution to (90).

During a large-scale epidemic, the probability that a randomly chosen vertex of degree k will become infected is given by

$$v_k = 1 - (1 - T + Tu)^k, \tag{99}$$

where u is the solution to (90). If he or she complies with an intervention that lowers the likelihood of transmission from infected contacts by a factor $\phi(0 \leqslant \phi \leqslant 1)$, then the probability of infection is reduced to

$$v_k = 1 - (1 - \phi T + \phi T u)^k. \tag{100}$$

References

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans, Dynamics and Control. Oxford University Press, Oxford.

Andersson, H., 1998. Limit theorems for a random graph epidemic model. Ann. Appl. Prob. 8, 1331–1349.

Avendano, M., Derkach, P., Swan, S. Clinical course and management of SARS in health care workers in Toronto: a case series. CMAJ 168, 1649–1660.

Ball, F., Mollison, D., Scalia-Tomba, G., 1997. Epidemics with two levels of mixing. Ann. Appl. Prob. 7, 46.

BC Stats, 2002. Labour force—employment, unemployment, related rates. http://www.bcstats.gov.bc.ca/data/lss/labour.htm

BC Stats, 2003. 2001 census profile of British Columbia's regions: Greater Vancouver regional district, www.bcstats.gov.bc.ca

Bozzette, S., Boer, R., Bhatnagar, V., Brower, J., Keeler, E., Morton, S., Stoto, M., 2003. A model for a smallpox-vaccination policy. N. Engl. J. Med. 348, 416–425.

Centre for Health Services and Policy Research, 2002. The British Columbia Health Atlas, second ed. http://www.chspr.ubc.ca/Research/healthatlas.htm#2ndEd

Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.-M., Thach, T.-Q., Chau, P., Chan, K.-P., Lam, T.-H., Tse, L.-Y., Tsang, T., Liu, S.-H., Kong, J.H.B., Lau, E.M.C., Ferguson, N.M., Anderson, R.M., 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 1.

Ferguson, N.M., Garnett, G.P., 2000. More realistic models of sexually transmitted disease transmission dynamics: sexual partnership networks, pair models, and moment closure. Sex Transm. Dis. 27, 600.

Hethcote, H., 2000. Mathematics of infectious diseases. SIAM Rev. 42, 599.

Jagers, P., 1975. Branching Processes with Biological Applications. Wiley, London.

Keeling, M.J., Woolhouse, M.E., May, R.M., Davies, G., Grenfell, B.T., 2003. Modelling vaccination strategies against foot-and-mouth disease. Nature 421, 136–142.

Kretzschmar, M., van Duynhoven, Y.T., Sverijnen, A.J., 1996. Modeling prevention strategies for gonorrhea and Chlamydia using stochastic network simulations. Am. J. Epidemiol. 144, 306–317.

Le, D., Bloom, S., Nguyen, Q., Maloney, S., Le, Q., Leitmeyer, K., Bach, H., Reynolds, M., Montgomery, J., Comer, J., Horby, P., Plant, A., 2003. Lack of SARS transmission among public hospital workers, Vietnam. Emerg. Infect. Dis. 10, 265–268.

Lloyd, A.L., May, R.M., 2001. Epidemiology. How viruses spread among computers and people. Science 292, 1316.

Longini, I.M., 1988. A mathematical model for predicting the geographic spread of new infectious agents. Math. Biosci. 90, 367.

McDonald, L., Simor, A., Su, I., Maloney, S., Ofner, M., Chen, K., Lando, J., McGeer, A., Lee, M., Jernigan, D., 2004. SARS in healthcare facilities, Toronto and Taiwan. Emerg. Infect. Dis. 10, 777–781.

Meyers, L.A., Newman, M.E.J., Martin, M., Schrag, S., 2003. Applying network theory to epidemics: control measures for *Mycoplasma pneumoniae* outbreaks. Emerg. Infect. Dis. 9, 204.

Meyers, L.A., Pourbohloul, B., Newman, M.E.J., Skowronski, D.M., Brunham, R., 2005. Network theory and SARS: predicting outbreak diversity. J. Theor. Biol. 232, 71–81.

Morris, M., 1995. Data driven network models for the spread of disease. In: Mollison, D. (Ed.), Epidemic Models: Their Structure and Relation to Data. Cambridge University Press, Cambridge, pp. 302–322.

Morris, M., Kretzschmar, M., 1997. Modeling prevention strategies for gonorrhea and Chlamydia using stochastic network simulations. AIDS 11, 641–648.

Newman, M.E.J., 2002. Spread of epidemic disease on networks. Phys. Rev. E 66 art. no.-016128.

Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2001. Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E 64, 026118.

Nicolosi, A., Correa Leite, M., Musicco, M., Arici, C., Gavazzeni, G., Lazzarin, A., 1994. The efficiency of male-to-female and female-to-male sexual transmission of the human immunodeficiency virus: a study of 730 stable couples. Italian study group on HIV heterosexual transmission. Epidemiology 5, 570–575.

Pourbohloul, B., Meyers, L.A., Skowronski, D.M., Krajden, M., Patrick, D.M., Brunham, R.C., 2005. Modeling control strategies of respiratory pathogens. Emerg. Infect. Dis. 11, 1249–1256.

Poutanen, S.M., Low, D.E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A.K., Skowronski, D.M., Salit, I., Simor, A.E., Slutsky, A.S., Doyle, P.W., Krajden, M., Petric, M., Brunham, R.C., McGeer, A.J., The National Microbiology Laboratory, and the Canadian Severe Acute Respiratory Syndrome Study Team, 2003. Identification of severe acute respiratory syndrome in Canada. N. Engl. J. Med. 348, 1995.

Sander, L.M., Warren, C.P., Sokolov, I.M., Simon, C.P., Koopman, J., 2002. Percolation on heterogeneous networks as a model for epidemics. Math. Biosci. 180, 293–305.

Sattenspiel, L., Simon, C.P., 1988. The spread and persistence of infectious diseases in structured populations. Math. Biosci. 90, 341.

Schwartz, N., Cohen, R., ben-Avraham, D., Barabasi, A.L., Havlin, S., 2002. Percolation in directed scale-free networks. Phys. Rev. E 66, 015104.

Skowronski, D., Astell, C., Brunham, R., Low, D., Petric, M., Roper, R., 2005. Severe acute respiratory syndrome: a year in review. Annu. Rev. Med. 56.

Statistics Canada, 2001. Household size, census metropolitan areas, http://www.statcan.ca/english/Pgdb/famil53i.htm

Tang, C., Wong, C., 2004. Factors influencing the wearing of facemasks to prevent the severe acute respiratory syndrome among adult Chinese in Hong Kong. Prev. Med. 39, 1187–1193.

Vancouver School Board, 2002. Vancouver School Board December 2002 Ready Reference, http://www.vsb.bc.ca/board/publications.htm

Varia, M., Wilson, S., Sarwal, S., McGeer, A., Gournis, E., Galanis, E., Henry, B., Team, H.O.I., 2003. Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada. CMAJ 169, 285–292.

World Health Organization, 2003. Consensus Document on the Epidemiology of Severe Acute Respiratory Syndrome (SARS). WHO Department of Communicable Disease Surveillance and Response, Geneva.

World Health Organization, 2004. Severe Acute Respiratory Syndrome (SARS). WHO Secretariat.

Xu, R., He, J., Evans, M., Peng, G., Field, H., Yu, D., Lee, C., Luo, H., Lin, W., Lin, P., Li, L., Liang, W., Lin, J., Schnur, A., 2004. Epidemiologic clues to SARS origin in China. Emerg. Infect. Dis. 10, 1030–1037.