CONTACT NETWORK EPIDEMIOLOGY: BOND PERCOLATION APPLIED TO INFECTIOUS DISEASE PREDICTION AND CONTROL

LAUREN ANCEL MEYERS

ABSTRACT. Mathematics has long been an important tool in infectious disease epidemiology. I will provide a brief overview of compartmental models, the dominant framework for modeling disease transmission, and then contact network epidemiology, a more powerful approach that applies bond percolation on random graphs to model the spread of infectious disease through heterogeneous populations. I will derive important epidemiological quantities using this approach and provide examples of its application to issues of public health.

BACKGROUND

Infectious diseases can have devastating impacts on human life and welfare. In the last three years, SARS, avian influenza, simian foamy virus, and monkeypox have jumped from animals into human populations. The uneven spread of SARS worldwide poignantly demonstrated that containment is possible but depends critically on appropriate and aggressive management. With the growing threats of newly emerging diseases and bioterrorism, strategies to rapidly and effectively control outbreaks are vital to public health.

Mathematics is an invaluable epidemiological tool. It allows public health officials to conduct virtual experiments that would be practically unfeasible or unethical. Controlled experiments to evaluate the efficacy of control strategies are impossible in practice, as we cannot intentionally introduce disease into populations or withhold potentially lifesaving interventions for the sake of scientific study. Mathematical models of disease transmission dynamics enable systematic evaluation of strategies such as vaccination and quarantine and thereby provide a way around this difficulty.

In the 18th century, Daniel Bernoulli—the son, nephew and brother of mathematicians Johann, Jacob and Nicolaus II Bernoulli, respectively—made one of the first great mathematical contributions to infectious disease control [BB04]. While formally trained in medicine, Bernoulli is known for his research in biomechanics, hydrodynamics, economics, and astronomy. He also played an important role in the eradication of smallpox from Europe, which was likely introduced there in the early 16th century and was endemic (maintained constantly) by the 18th century. Variolation is an inoculation technique whereby a scab or pus from an individual with a

---

mild smallpox infection is introduced into the nose or mouth of healthy individuals. This practice began as early as 1000 AD in China and India and was introduced into England in 1717, where it was initially controversial. While variolation reduced the mortality probability of infected individuals from 30% to 1% [Sma02], there was a small chance that the procedure would lead to death from a full-blown case of smallpox.

Bernoulli developed a mathematical model with which he argued that the gain from variolation in life expectancy through the eradication of smallpox far outweighed associated risks [BB04, DH02]. Assuming that all individuals had a one in n chance of catching smallpox and a one in m chance of dying from an infection, he derived the following equation for the change in the number of currently naïve (never been infected) individuals in a specific age cohort during a small increment of time:

$$(1) \qquad -\mathrm{d}s = \frac{s\,\mathrm{d}x}{n} + \frac{s}{\xi}\left((-\mathrm{d}\xi) - \frac{s\,\mathrm{d}x}{mn}\right)$$

where dx is the change in the age of the individuals in the cohort, s and ds are the number of currently naïve individuals and the change in that number, respectively, and ξ and $d\xi$ are the total number of individuals in the cohort and the change in that number, respectively. On the right side of the equation, the first term is the number of new infections and the second term is the loss of susceptible individuals through death from other causes. Bernoulli integrated equation (1) and assumed that each cohort is born entirely susceptible (that is, $s = \xi$ when $x = 0$) to find the expected fraction of susceptible individuals in a cohort of age x. This fraction is given by

$$(2) \qquad \frac{s}{\xi} = \frac{m}{(m-1)\,\mathrm{e}^{\frac{x}{n}} + 1}.$$

Bernoulli assumed that the risk of catching smallpox was 12.5% (one in eight) in a given year across all age classes and that the mortality rate was 12.5% (one in eight) for all infected individuals. Using overall survivorship estimates calculated by Edmund Halley (of comet fame), he then used equation (2) to predict the mortality rates in every age class in a steady-state population with a birth class of size 1300.

Inoculation via variolation of all newborns would confer widespread immunity, yet entail some mortality due to variolation-induced smallpox. Bernoulli compared the annual mortality rates and average life expectancy predicted by his model to those predicted assuming universal inoculation and found that variolation saves lives even if the mortality rate associated with variolation is quite high (with his parameters, as high as 10.6%).

Bernoulli's calculations clarified the benefits of widespread inoculation even when there are significant risks. England began widely administering variolation in the 1750's, and upon the development of the smallpox vaccine in 1796, mandated the inoculation of all infants. Thanks to these efforts, smallpox was eradicated from England by the end of the 19th century.

Since Bernoulli, mathematicians and statisticians have offered many practical insights into infectious disease control. Notably, the English statistician William Farr analyzed the spatial distribution of cholera cases and thereby provided the first solid evidence that the disease spread via water rather than air [PVJ98].

Mathematical epidemiology exploded in the 20th century following the introduction of an intuitive and tractable framework. Between 1906 and 1927, the mass-action principle was introduced [Ham06] and ultimately formalized in a deterministic model of disease transmission now attributed to Kermack and McKendrick [KM27]. In chemistry, the mass-action law states that the rate of a chemical reaction is proportional to the product of the concentrations of the reacting substances. In epidemiology, the mass-action assumption states that the number of new cases of disease in a time interval is proportional to the product of numbers of infected and susceptible hosts in the previous time interval. Within a decade, Reed and Frost introduced the first stochastic version of this model, the chain-binomial, which assumes that a disease spreads in discrete generations [Abb52, Bai75]. This model derives a probability law for the next generation from that of the present generation. More recently, Anderson and May among others have extended these efforts into a flexible approach, known as compartmental modeling, for predicting the transmission of a wide range of diseases on multiple scales [AM91].

We will discuss this framework as well as some of its practical applications and limitations below. This will set the stage for the introduction of contact network epidemiology, a new analytical approach that overcomes a major limitation of the mass-action assumption.

Compartmental SIR models

Compartmental models subdivide host populations by disease status. A simple and widely used example is the SIR model that tracks the movement of hosts among three states: susceptible (S), infected (I), and recovered (R) (Figure 1) [KM27]. These models assume that upon infection, hosts are immediately infectious and remain infectious until they recover. Infected hosts are assumed to have potentially disease-causing contacts with *random* individuals from the population according to a Poisson process that yields an average contact rate of β per unit time. Disease transmission occurs if and only if the individual at the receiving end of the contact is susceptible. Therein lies the mass-action assumption.

Infectious hosts leave the infectious state at an average rate ν either by recovering and becoming immune or by dying. Thus the recovered class is a catchall for hosts that have been previously infected and are no longer infected or susceptible. In the limit of a large host population, this process is modeled by the following coupled nonlinear differential equations:

(3)
$$\frac{dS}{dt} = -\beta IS,$$
$$\frac{dI}{dt} = \beta IS - \nu I,$$
$$\frac{dR}{dt} = \nu I,$$

where $S(t)$, $I(t)$, and $R(t)$ are the numbers of susceptible, infected, and recovered hosts, respectively. Because the model ignores the birth and death of susceptibles, the total population size $N = S + I + R$ is static, and therefore the third equation is unnecessary. These equations apply to rapidly spreading diseases like measles and influenza that confer immunity extending beyond the typical length of an epidemic. The model can be easily adapted to consider the loss of immunity as well as birth and death dynamics.
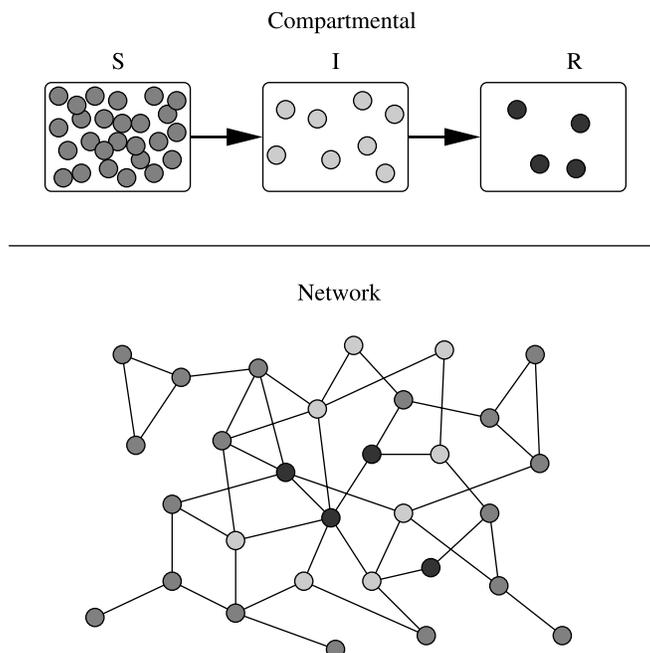
Compartmental

Network

FIGURE 1. Compartmental and contact network models. Mass-action models assume that all individuals in a group are equally likely to become infected, while contact network epidemiology considers diverse contact patterns that underlie disease transmission. The disease spreads along the arrows (top) and the edges (bottom). (S=susceptible, I=infected, and R=recovered.)

The basic reproductive rate. One of the touchstones of epidemiology is the *basic reproductive rate* of a disease: the number of secondary infections produced by a single infected host in an entirely susceptible population. This quantity indicates the initial growth rate for the infected class and the potential for a large-scale epidemic. In the model given by equations (3), the per capita increase of infected individuals is given by

$$(4) \qquad \frac{1}{I}\frac{dI}{dt} = \beta S - \nu.$$

The number of infected individuals increases by the product of the disease-causing contact rate β and number of susceptibles S and decreases by the combined recovery and mortality rate (henceforth *removal rate*) ν, which has units of $\frac{1}{t}$. The reciprocal of the removal rate, $\frac{1}{\nu}$, is the average time interval during which an infected individual remains contagious. The number of secondary cases infected per unit time is βS, which yields a basic reproductive rate of

$$(5) \qquad R_0 = \frac{\beta S}{\nu}.$$

If $R_0 > 1$, then each infected host will transmit disease to at least one other host during the infectious period, and the model predicts that disease will spread through the population. If not, then the disease is expected to fizzle out before reaching

a substantial fraction of the population. Thus $R_0 = 1$ is a critical epidemiological value. In other words, pathogens with high levels of contagion and low recovery and mortality rates will pose the greatest threat.

Herd immunity. The immunization of a single host not only protects that host but also indirectly protects others against the possibility of disease transmission from the immunized host. If a sufficient fraction of a population is immunized, then an epidemic may be averted altogether. The protection of an entire population via the immunity of a fraction of the population is called *herd immunity*.

Equation (5) can be rearranged to find the minimum size of a susceptible population necessary for an epidemic to occur. Assuming that $R_0 = 1$, this threshold is given by

$$(6) \qquad S_T = \frac{\nu}{\beta}.$$

A pathogen will go extinct if the size of the susceptible population is less than this threshold ($S < S_T$). If the population size is above this threshold, then we can rewrite the basic reproductive rate as

$$(7) \qquad R_0 = \frac{S}{S_T}.$$

Immunization reduces the size of the susceptible class and thus leads to a smaller basic reproductive rate of the pathogen. In particular, immunizing a fraction p of a population reduces R_0 to

$$(8) \qquad R_0^i = \frac{(1-p)\, S}{S_T} = (1-p)\, R_0.$$

Immunization will successfully eradicate the disease if it causes the basic reproductive rate to drop below one. Thus the critical immunization rate p_c is

$$(9) \qquad p_c = 1 - \frac{1}{R_0}.$$

Extensions of this basic model have been used to predict the minimum coverage necessary to drive specific diseases to extinction. For example, measles and whooping cough—two of the most contagious diseases—are thought to require 90-95% coverage, chicken pox and mumps 85-90% coverage, polio and scarlet fever 82-97% coverage, and smallpox 70-80% coverage [AM91].

Limitations of the mass-action assumption: The example of SARS.
Shortly after severe acute respiratory syndrome (SARS) was first recognized outside of Asia, epidemiologists estimated its basic reproductive rate (R_0) to be between 2.2 and 3.6 for this virus [LCC03, RFD03]. These estimates are well above one and similar to rates measured for new subtypes of influenza [Het00]. Despite this high estimate and worldwide susceptibility to SARS, the disease did not spark a global pandemic.

The discrepancy between the high R_0 estimates and the limited spread of SARS might be explained by effective public health intervention that reduced the basic reproductive rate of the disease. Consider, however, the transmission of SARS in China during its initial four months of spread before the implementation of extensive public health measures. Case counts were much less than expected during this period [XHE04], as suggested by a simple calculation. The expected total number of cases of a disease is predicted to increase by a factor of R_0 for every generation of

disease transmission, where a generation is the average time between an individual becoming infected and their infecting others. The average generation time (γ) for SARS was estimated to be 9.7 ± 0.3 days [MPN05]. The cumulative number of SARS cases after D days of transmission is predicted to be approximately

$$(10) \qquad \sum_{i=0}^{D/\gamma} (R_0)^i = \frac{1 - R_0^{D/\gamma+1}}{1 - R_0}.$$

For R_0 ranging between 2.2 and 3.6, this then suggests that the first four months of SARS spreading in China should have produced somewhere between approximately 30,000 and 10 million cases [MPN05]. China ultimately reported only 782 cases during this initial outbreak [Org03], which, by equation (10), suggests that the reproductive rate of SARS was actually closer to 1.6.

Why do the initial estimates of R_0 seem incompatible with the observed epidemiology in China? The basic reproductive rate has two critical inputs: (1) intrinsic properties of the pathogen that determine the transmission efficiency per contact and the duration of the infectious period, and (2) the patterns of contacts between infected and susceptible hosts in the population. While the first factor may be fairly uniform across outbreaks, the second may be quite context dependent, varying both within and among populations. The problem with the SARS estimates stems from the mass-action assumption of compartmental models—that all susceptible individuals are equally likely to become infected. When this assumption does not hold, the models may yield inaccurate estimates or estimates that do not apply to all populations. The R_0 estimates for SARS [LCC03, RFD03] were based largely on outbreak data from a hospital and a crowded apartment building, with anomalously high rates of close contacts among individuals. It may thus be inappropriate to extrapolate estimates for R_0 from these specific settings to the population at large. Contact rates in the general community may be much lower and, therefore, so may be the rate at which SARS spreads.

The transmission efficiency of SARS varied considerably. A few individuals were responsible for a large proportion of disease transmission [LCH03, BMT03, DGL03]. In contrast to the mass-action assumption of standard compartmental models, contact patterns may vary within a community. Consider two scenarios: a community in which all individuals have approximately the same number of contacts and a community in which a very small number of individuals have enormous numbers of contacts while all other individuals have only one or very few contacts. The basic reproductive rate of disease (R_0) can be identical for the two communities, while the resulting epidemiology will differ significantly.

While the mass-action assumption laid the groundwork for major advances in epidemiological theory, it may be inappropriate when contact patterns are heterogeneous. To overcome this limitation, mathematical epidemiologists have developed several methods to explicitly consider heterogeneity in contact patterns including more complex deterministic and stochastic compartmental models with multiple demographic groups [Bai75, HY84], branching process models [Bec77, FKG03], dyad models [KRM97, FG00], Reed-Frost chain-binomial models [LP89], and individual-based models [KG99, Dur99, SWS02, BRO02, VdPvVdV98, CHECC03, EGK04].

Here we will consider a recent addition to this toolkit, *contact network epidemiology*, which is an analytical framework that explicitly and intuitively captures the diverse interactions that underlie the spread of diseases (Figure 1) [MPN05, KWM03, MNP06, DdJM98, New02, SS88, BMST97, Mor95, Lon88, LM01, MNMS03].

Contact network epidemiology

The methods of contact network epidemiology can be divided into three steps. First we attempt to build a realistic network (graph) model of the contact patterns at an appropriate temporal and spatial scale. Second, we mathematically predict the spread of disease through the population based on intrinsic features of the pathogen and structural properties of the network. Third, we manipulate the network to model control strategies and analyze the epidemiological impact of such manipulations. We will now discuss each of these steps with illustrative examples.

The contact networks. A contact network model captures the patterns of interactions that can lead to the transmission of an infectious disease. In a contact network, each person (or location) translates into a *vertex*, and contacts among people (or locations) translate into *edges* that connect appropriate vertices. For example, one might model the contacts between individuals in a hospital or city that might lead to respiratory disease transmission [MPN05, EGK04, MNP06, MNMS03], the contacts between different geographical regions via human travel patterns that might lead to long-range transmission, or the sexual interactions within a high school that might lead to sexually transmitted disease transmission [Rot01, RST97].

The number of edges emanating from a vertex is called the *degree* of the vertex and indicates the number of possible contacts that can lead to disease transmission to or from an individual. The distribution of the number of such contacts within a population (the *degree distribution*) is fundamental to the ability of disease to spread through the population. The mass-action assumption of compartmental models is tantamount to assuming that the underlying contact patterns form a random graph with a Poisson degree distribution. If a network departs significantly from this ideal structure, then the traditional modeling approach may be invalid.

The contact (or social) network is a hot concept across many disciplines, including sociology, epidemiology, biology, computer science and physics [AO04]. Researchers look for universal properties and have paid special attention to small-world networks, characterized by high levels of both local clustering and global connectivity [Wat99]; and scale free networks, characterized by degree distributions that follow a power law distribution with a small fraction of very highly connected hubs [BA99]. Several epidemiological-relevant contact networks, including sexual contact networks and the Internet, for example, have been characterized as scale free [LEA01, LEA03, PSV01].

Realistic contact networks, however, do not always fall into one of these well-studied families of networks [MNMS03, MPN05, MNP06]. Some have more complex structures, for example, those depicted in Figure 2. Bipartite networks, in which there are two types of nodes, have been used to represent asymmetric probabilities of transmission between caregivers and patients in a medical facility [MNMS03]. Semi-directed networks, in which some contacts are reciprocal and others are unidirectional, have been used to capture situations in which a person may infect another person but the converse is not true [MNP06]. This situation may arise, for example, when infected individuals seek medical treatment during an outbreak.
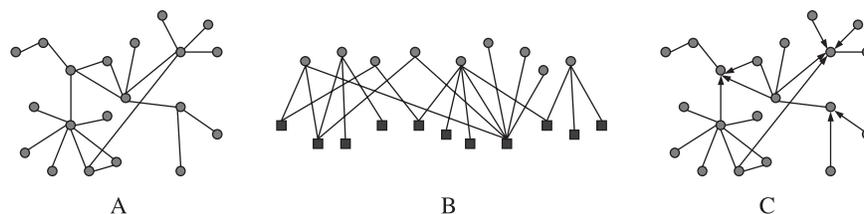
FIGURE 2. Contact networks. (A) Undirected network; (B) Bipartite network; and (C) Semi-directed network. Figure from [MNP06].

Suppose individual A is normally healthy and thus has no reason to go to the hospital until he or she becomes infected. At that point, individual A may come into contact and potentially spread disease to caregivers at the hospital. In contrast, if a caregiver at the hospital acquired the disease while individual A remained healthy, then there would be no opportunity for transmission in the opposite direction. This asymmetry can be modeled by directed edges pointing from individual A to health care workers. As described next, the mathematical methods of contact network epidemiology can accommodate such complex random networks with arbitrary degree distributions.

Predicting disease dynamics. Imagine that an infectious disease first appears at a randomly chosen vertex in a contact network (epidemiologically speaking, that vertex represents *patient zero*). Disease will propagate through the network as described for the compartmental models, except that the Poisson distribution of contacts is replaced by the structure of the contact network. The initial vertex will remain infected and infectious for some period of time, during which it has the potential to transmit disease to each of its contacts. The secondary cases likewise can transmit disease to their contacts during their infectious period, and so on. This process resembles bond percolation and can be analyzed using percolation models from statistical physics [New02, MPN05, MNP06, MNMS03, SWS02, PMS05]. This approach was initially suggested by Grassberger [Gra83] and has been extended into a flexible framework for infectious disease prediction by Newman and colleagues [New02, MPN05, MNP06, MNMS03, FBMBss]. In what follows, we will review Newman's derivations of fundamental epidemiological quantities for an undirected random network with an arbitrary degree distribution and some practical corollaries.

The percolation of disease through a network depends on both the level of contagion and the structure of the contact network. Following Newman [New02], every edge in a network has a per unit time probability of disease transmission associated with it (r_{ij}), that is, the probability that vertex i, if infected, will transmit disease to vertex j in a given time increment. Assuming discrete time steps, if vertex i is infectious for τ time steps, then the probability that j will be infected by i is $T_{ij} = 1 - (1 - r_{ij})^{\tau}$. For continuous time, $1 - T_{ij} = \lim_{\delta t \to 0} (1 - r_{ij}\delta t)^{r/\delta t} = \mathrm{e}^{-r_{ij}\tau}$, and thus $T_{ij} = 1 - e^{-r_{ij}\tau}$. The quantity r_{ij} summarizes core aspects of disease transmission, including the likelihood that a contact will lead to transmission and

individual susceptibility and will therefore vary across individuals. If r_{ij} is assumed to be an independent identically distributed (iid) random variable chosen from a distribution $P(r)$, then T_{ij} is also an iid random variable. Therefore the spread of disease will depend only on the mean probability of transmission between individuals (henceforth, the average transmissibility), which is given by

$$(11) \qquad T = \langle T_{ij} \rangle = 1 - \int_0^\infty Q(r)\,dr$$

where $Q(r) = 1 - P(r)(1-r)^\tau$ or $Q(r) = 1 - P(r)e^{-r\tau}$, for discrete or continuous time, respectively.

Probability generating functions. To predict the fate of an outbreak, we use probability generating functions (pgf's), quantities that describe probability distributions, and here summarize useful information about the structure of the contact network. The pgf for the degree distribution of a network is

$$(12) \qquad G_0(x) = \sum_{k=1}^\infty p_k x^k$$

where p_k is the relative frequency of vertices of degree k in the network. The average degree equals the derivative of this function at $x = 1$, that is, $\langle k \rangle = \sum_{k=1}^\infty k p_k$.

If we choose a random edge and follow it to one of its vertices, then the number of remaining edges connected to the vertex is called the *excess degree* of the vertex. The higher the degree of a vertex, the more likely it is to lie at the end of a randomly chosen edge. In particular, the likelihood of reaching a vertex with degree k, and thus with excess degree $k-1$, will be proportional to k. Thus the probability that a vertex at the end of a random edge has excess degree $k-1$ is $\frac{kp_k}{\langle k \rangle}$. This yields a generating function for the excess degree of a vertex of

$$(13) \qquad G_1(x) = \frac{\sum_{k=1}^\infty k p_k x^{k-1}}{\sum_{k=1}^\infty k p_k}.$$

and an average excess degree of

$$\langle k_e \rangle = \frac{\sum_{k=1}^\infty k(k-1)p_k}{\sum_{k=1}^\infty k p_k} = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1.$$

When disease is introduced into a network, it will traverse some but not all of the edges according to the average transmissibility T. The edges that are infected during an epidemic are called *occupied*. Once the disease has run its course, the cluster of vertices connected to the first infected vertex by a continuous chain of occupied edges is exactly the outbreak. Ultimately we will characterize the size and distribution of this occupied cluster. We begin by deriving the pgf for the distribution of occupied edges attached to a randomly chosen vertex as a function of the average transmissibility T. The probability that a vertex has m of its k edges occupied is simply $\begin{pmatrix} k \\ m \end{pmatrix} T^m (1-T)^{k-m}$, which leads to a probability generating

function for the occupied degree of a vertex of

$$(14) \quad G_0\left(x;T\right) = \sum_{m=0}^{\infty}\sum_{k=m}^{\infty} p_k \left(\begin{array}{c} k \\ m \end{array}\right) T^m \left(1-T\right)^{k-m} x^m = G_0\left(1 + (x-1)T\right).$$

Analogously, the pgf for the excess occupied degree, that is, the number of occupied edges emanating from a vertex reached by following a randomly chosen edge, is given by

$$(15) \qquad\qquad G_1\left(x;T\right) = G_1\left(1 + (x-1)T\right).$$

Predicting the fate of a small outbreak. In general, percolation theory describes the behavior of connected groups of vertices in a random graph and thus can be applied 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, 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.

First we will consider Newman's derivation of the epidemic threshold and the expected size of small outbreaks below the threshold [New02]. These calculations assume that mildly contagious diseases spread via branching, yielding short transmission chains that do not loop back on themselves. Later, we relax this assumption and consider diseases that are above the epidemic threshold.

Let s denote the number of vertices contained in a small outbreak that begins at a randomly selected vertex and let $H_0(x;T)$ be the generating function for the distribution of outbreak sizes. Then

$$(16) \qquad\qquad H_0(x;T) = \sum_s P_s(T)x^s$$

where $P_s(T)$ is the probability that a single initial case sparks an outbreak of size s at the specified average transmissibility T. Let $H_1(x;T)$ be the generating function for the size of the cluster of connected vertices at the end of a randomly chosen edge.

To solve for the average value of s, we consider an outbreak that originates with a transmission event along a randomly chosen edge. The set of vertices reached by occupied edges can be represented in graphical form as in Figure 3. There are many possible outcomes: the disease does not spread along the edge, it spreads along the edge but no further, or it spreads along the edge and then subsequently along one or more additional edges emanating from the destination vertex. This is captured in a recursive equation

$$(17) \qquad\qquad H_1\left(x;T\right) = xG_1\left(H_1\left(x;T\right);T\right).$$

This is roughly interpreted to mean that the size of a cluster proceeding from a randomly chosen edge E is the equal to the sum of the sizes of the clusters at the end of each occupied edge emanating from the vertex V at the end E plus one for the vertex V itself. Likewise, the cluster emanating from a random vertex is generated by

$$(18) \qquad\qquad H_0\left(x;T\right) = xG_0\left(H_1\left(x;T\right);T\right).$$
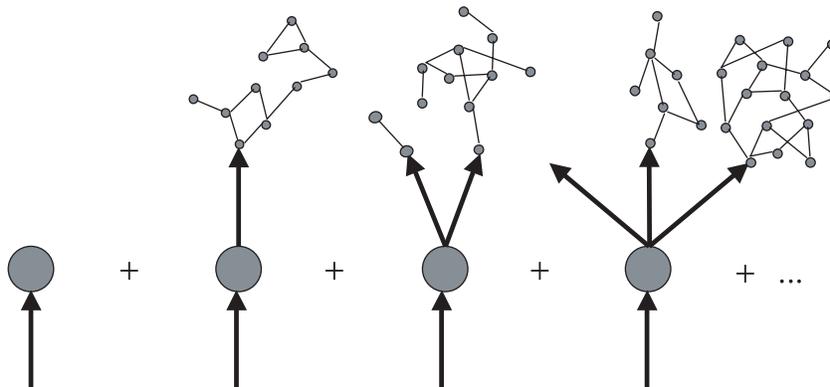
FIGURE 3. Future transmission diagram. When disease reaches an edge, we can consider all possible patterns of future transmission. The disease may not spread along the edge in the first place, it may spread along the edge but no further, it may spread along the original edge and then subsequently along another edge, it may spread along the original edge and then subsequently along two edges, etc. We can construct recursive equations to consider all possible outcomes.

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

(19)
$$\langle s \rangle = \sum_s s P_s(T) = H_0'(1;T) = 1 + G_0'(1;T)H_1'(1;T)$$
$$= 1 + \frac{TG_0'(1)}{1 - TG_1'(1)} = 1 + \frac{T\langle k \rangle}{1 - T\left(\langle k^2 \rangle / \langle k \rangle - 1\right)}$$

where the prime denotes differentiation with respect to the first variable. Note that for any normalized generating function $f(x)$, $f(1) = 1$. The expression for $\langle s \rangle$ diverges when the denominator in equation (19) is zero and predicts the expected size of the outbreak only when the denominator is greater than zero. Thus the equation

(20)
$$TG_1'(1) = 1$$

marks the phase transition at which the size of an outbreak first becomes extensive. Hence an epidemic is possible only when the average transmissibility of a disease is greater than the critical transmissibility

(21)
$$T_c = \frac{1}{G_1'(1)} = \frac{\sum_k k p_k}{\sum_k k(k-1)p_k} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle},$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the mean degree and mean square degree (respectively) of the network. We call T_c the *epidemic threshold*.

The basic reproductive rate. Recall that the basic reproductive rate is the number of secondary infections caused by a single infected host in a completely

naïve population. In the contact network framework, this is simply the average number of occupied edges emanating from a vertex, that is,

$$(22) \qquad R_0 = G_1'\left(1;T\right) = TG_1'\left(1\right) = T\left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}\right).$$

Recall that T_c is the critical transmissibility value above which population is vulnerable to large-scale epidemics (but is not guaranteed to experience an epidemic) and below which only small local outbreaks occur. If the transmissibility of a disease equals the epidemic threshold ($T = T_c$), then $R_0 = 1$.

Notice that the basic reproductive rate depends explicitly on the structure of the network (on $\langle k \rangle$ and $\langle k^2 \rangle$). A single pathogen may therefore have very different transmission dynamics depending on the population through which it spreads. If two networks have the same mean degree, $\langle k \rangle$, then the one with the larger variance in degree, $\langle k^2 \rangle - \langle k \rangle^2$, will be more vulnerable to the spread of disease. Estimates of R_0 that assume a mass-action model may therefore be invalid for populations with non-Poisson contact patterns and, in particular, underestimate the actual growth rate of the disease in highly heterogeneous networks.

Probability and size of a large-scale epidemic. When the transmissibility of a disease lies above the epidemic threshold, then equation (19) no longer applies. Disease transmission is so frequent that transmission chains are likely to loop back upon themselves and, therefore, the assumption underlying the calculations (depicted in Figure 3) no longer holds. When a disease is above the epidemic threshold, an epidemic can occur, and we can calculate two quantities: the probability that a large-scale epidemic occurs and the expected fraction of individuals infected in such an epidemic. In an undirected network, these quantities are equal to each other and to the fraction of vertices from which an extensive numbers of others can be reached by following occupied edges. In the language of percolation, this is the giant component defined by occupied edges.

The probability of a full-blown epidemic, S, is derived by first calculating the likelihood that a single infection will lead to only a small outbreak instead of a full-blown epidemic, and then subtracting that value from one. Recall that $H_0(x;T)$ is the generating function for the size of small outbreaks. Therefore $H_0(1;T)$ is the total probability that a randomly chosen initially infected vertex will lead to a finite-sized outbreak. The probability of a large-scale epidemic is then given by

$$(23) \qquad S = 1 - H_0\left(1;T\right) = 1 - G_0\left(u;T\right)$$

where $u = H_1\left(1;T\right)$. Thus u is the solution to the equation

$$(24) \qquad u = G_1\left(u;T\right).$$

In terms of the degree distribution, the probability of a large-scale epidemic and the expected fraction of the network infected during such an epidemic is

$$(25) \qquad S = 1 - \sum_k p_k \left(1 + (u-1)\,T\right)^k$$

where u is the solution to the self-consistency equation

$$(26) \qquad u = \frac{\sum_k k p_k \left(1 + (u-1)\,T\right)^{k-1}}{\sum_k k p_k}.$$

We use numerical root finding methods to solve for u.

Other useful epidemiological quantities. We have recently extended Newman's results [New02] to provide insight into other epidemiological processes. In particular, we have derived the probability of becoming infected and sparking an infection as a function of the degree of a vertex, the probability of an epidemic starting from an outbreak that is already under way, and the residual structure of a network after an epidemic has run its course.

The probability that an individual will spark an epidemic [MPN05]. The probability ε_k that an individual (patient zero) with degree k will start an epidemic is equal to the probability that transmission along at least one of the k edges emanating from that vertex will lead to an epidemic. For any of its k edges, the probability that the disease does not get transmitted along the edge is $1 - T$, and the probability that disease is transmitted to the attached vertex but does not proceed into a full-blown epidemic is Tu. Thus

$$(27) \qquad \varepsilon_k = 1 - (1 - T + Tu)^k.$$

The probability that a disease cluster will spark an epidemic [MPN05]. The probability that an outbreak with N cases will spark a large-scale epidemic is given by $1 - \prod_{i=1}^{N} (1 - \varepsilon_{k_i})$ where k_i is the degree of individual i in the original outbreak. This is simply one minus the probability that not one of the N infected individuals starts an epidemic. Suppose now that we know the number of cases in the initial outbreak but not the degrees of each of those individuals. Then we can estimate the probability of an epidemic similarly, replacing each of the $(1 - \varepsilon_{k_i})$'s with the probability that a typical infected individual does not start an epidemic. The excess degree pgf gives the number of edges along which an infected individual can initiate an epidemic, and $1 - T + Tu$ is the probability that one of those edges will lead to an epidemic. Combining the two, we find that the probability that none of the edges on a typical infected individual will lead to an epidemic is

$$\left(\frac{\sum_{k=1}^{\infty} k p_k (1 - T + Tu)^{k-1}}{\sum_{k=1}^{\infty} k p_k} \right),$$

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

$$1 - \left(\frac{\sum_{k=1}^{\infty} k p_k (1 - T + Tu)^{k-1}}{\sum_{k=1}^{\infty} k p_k} \right)^N.$$

Individual risk of infection [MPN05]. The probability ν_k that an individual with degree k will be infected during a large-scale epidemic is equal to one minus the probability that disease is not transmitted along any of the k edges leading to that individual. The probability that a contact does not transmit the disease is equal to the probability that the contact was never infected, u, plus the probability that the contact was infected at some point but did not transmit the disease to the

individual in question, $(1-u)(1-T)$. A randomly chosen vertex of degree k will therefore become infected during an epidemic with probability

$$(28) \qquad \nu_k = \varepsilon_k = 1 - (1 - T + Tu)^k.$$

Frailty and interference [FBMBss]. If hosts are immunized following infection, an epidemic will change the structure of the *epidemiologically active network* (the remaining susceptible nodes and the edges that connect them). We characterize the structural evolution of a network due to an epidemic in terms of *frailty*, the degree to which highly connected individuals are more vulnerable to infection; and *interference*, the extent to which the epidemic cuts off connectivity among the susceptible population that remains following an epidemic. For a vertex that never becomes infected during an epidemic, we can distinguish between its original degree k and its degree in the residual network consisting of all nodes that remain uninfected by the epidemic, k_r. To understand the structural evolution of the network we derive two new network statistics: the mean *original* degree of individuals remaining in the residual network $\langle k \rangle_r$ and the mean *residual* degree of the individuals remaining in the residual network $\langle k_r \rangle_r$.

Recall that ν_k is the probability that a randomly chosen vertex of degree k will become infected in an outbreak. The fraction of individuals of degree k who remain uninfected throughout an epidemic is given by $r_k = \frac{p_k(1-v_k)}{\sum_j p_j(1-v_j)}$. Therefore the mean

original degree in the residual network is given by

$$(29) \qquad \langle k \rangle_r = \sum_k k r_k = \frac{\sum k p_k (1 - v_k)}{\sum p_k (1 - v_k)}.$$

To calculate the residual degree after an epidemic, we divide the network into the vertices that become infected and those that do not and then calculate the fraction of edges that connect individuals within the uninfected group. We call the two ends of an edge its *stubs*. A vertex with degree k will have exactly k stubs attached to it, and the total number of stubs in the network is $N \sum k p_k$ where N is the number of vertices. For each of the approximately $N \sum p_k v_k$ vertices in the infected group, infection of that vertex necessarily occurred along an edge with both its origin and destination stub attached to vertices in the infected group. For simplicity, we ignore the first infection, which did not occur along an edge within the population. The total number of stubs in the network excluding those along which infections took place is then $N \sum p_k (k - 2v_k)$. In other words, this quantity is the total number of uninfected stubs.

Now suppose we randomly choose an edge along which disease was not transmitted during an epidemic. We can calculate the probability that the node at the end of the edge was infected during the epidemic. If the degree of this node is k, then this equals the probability that it was not infected along any of its other $k-1$ edges. Recall that the probability that a contact does not transmit the disease is equal to the probability that the contact was never infected, u, plus the probability that the contact was infected at some point but did not transmit the disease to the individual in question, $(1-u)(1-T)$. Thus the probability that the node of degree k at the end of a randomly chosen uninfected edge was infected is $\omega_k = 1 - (1 - T + Tu)^{k-1}$.

We use this quantity to calculate the following probability that a randomly chosen uninfected stub attaches to an uninfected node: $\frac{\sum k p_k (1 - \omega_k)}{\sum p_k (k - 2v_k)}$. If instead

we randomly choose an uninfected edge rather than an uninfected stub, then the probability that it connects two uninfected nodes is given by the square of that quantity, $\left(\frac{\sum kp_k(1-\omega_k)}{\sum p_k(k-2v_k)}\right)^2$. To then calculate the average residual degree in the residual network, we multiply this probability by the total number of stubs and then divide by the number of vertices remaining in the residual network,

(30)
$$\langle k_r \rangle_r = \left(\frac{\sum kp_k(1-\omega_k)}{\sum p_k(k-2v_k)}\right)^2 \left(\frac{N\sum p_k(k-2v_k)}{N\sum p_k(1-v_k)}\right) = \frac{\left(\sum kp_k(1-\omega_k)\right)^2}{\sum p_k(k-2v_k)\sum p_k(1-v_k)}.$$

Taking a slightly different approach, Newman derived a similar quantity for the average residual degree [New05].

We next calculate frailty, the extent to which high degree vertices are preferentially infected during an epidemic. In particular, frailty is equal to the difference between the mean degree in the original network and the mean original degree in the residual network, scaled by the mean original degree,

(31)
$$\phi = \frac{\langle k \rangle - \langle k \rangle_r}{\langle k \rangle}.$$

Finally we calculate interference, the extent to which the epidemic disconnected the remaining susceptible population. Interference is calculated as the difference between the mean original degree in the residual network and the mean residual degree in the residual network, again scaled by the mean original degree,

(32)
$$\theta = \frac{\langle k \rangle_r - \langle k_r \rangle_r}{\langle k \rangle}.$$

Epidemiological dynamics on random networks. All of the quantities above pertain to the final outcome of an outbreak or epidemic. Volz has recently developed a system of nonlinear differential equations to model the dynamical progression of a disease spreading through a random network with arbitrary degree distributions [Volss]. His model considers the state of each edge and each stub (one end of an edge) in the network. An edge is considered *occupied* if it has ever been a conduit for disease transmission, *refractory* if it is connected to a recovered vertex and is not occupied, and *susceptible* if it is neither occupied nor refractory. The state of the stub depends on the state of its edge and on which end of the edge it occupies. The four equations of the model track the changing distribution of edge and stub states as disease percolates through the network. This model provides important insight into the interaction between population structure and epidemiological dynamics and will be an important tool for optimizing the timing and targets of control measures.

Predictions on more complex contact networks. Newman extended the analysis described above to random networks in which transmission rate is correlated with the degree of either the infecting vertex or the vertex becoming infected [New02]. We have derived the epidemiological quantities described above for bipartite [MNMS03] and semi-directed contact networks [MNP06]. Here we briefly discuss some interesting features of disease transmission on semi-directed networks.

In a semi-directed network, each vertex has an *undirected* degree representing the number of undirected edges attached to the vertex as well as both an *in-degree* and an *out-degree* representing the number of directed edges pointing towards or

away from the vertex, respectively. The undirected-degree and in-degree indicate the numbers of contacts that can spread disease to the individual and thus is related to the epidemiological vulnerability of the individual. The undirected-degree and out-degree indicate the numbers of contacts that may be infected by that individual should he or she become infected and thus is related to the potential epidemiological impact of an individual. The *semi-directed degree distribution* is the joint probability distribution p_{jkm} that a vertex has j incoming edges, k outgoing edges, and m undirected edges.

Our derivations in [MNP06] reveal that semi-directed networks are more complicated than undirected networks in two important respects. First, there can be two different distributions of transmission rates: one for the directed edges and one for the undirected edges. When these distributions differ, the epidemic threshold is no longer a single value but a line dividing the two-dimensional 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.

Second, recall that, in an undirected network, the probability of an epidemic and the expected fraction of the network infected during an epidemic are equal. In a semi-directed network, however, when the in-degree and out-degree distributions differ, then so do the probability of an epidemic and the expected incidence should one occur. These quantities are equivalent to the fraction of vertices from which an extensive number 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 theory, the set of vertices from which one can reach most other vertices consists of the giant strongly connected component (GSCC)
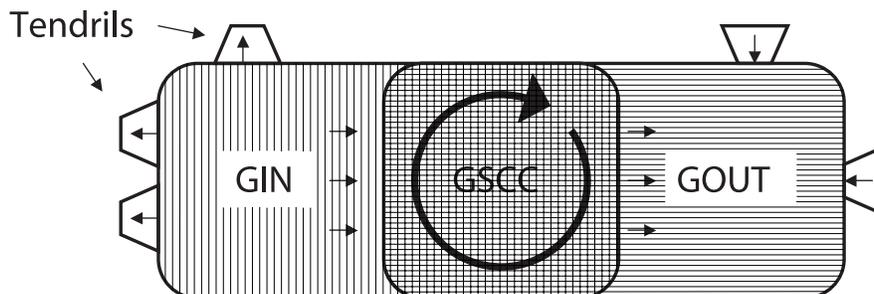

FIGURE 4. Structure of a semi-directed network. The largest set of vertices in which you can move between any two by following edges in the correct direction is the *giant strongly connected component* (GSCC). The set of vertices not contained in the GSCC that can be reached by following edges in the 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. Figure from [MNP06].

and the giant in-component (GIN), whereas the set of vertices that can be reached from most others consists of the GSCC and the giant out-component (GOUT). Figure 4 provides a diagram of a typical semi-directed network. The probability of a large-scale epidemic corresponds to the relative size of the region shaded in vertical lines, and the expected fraction of the population that will become infected during such an epidemic corresponds to the relative size of the region shaded in horizontal lines.

Evaluating control strategies

A primary public health goal is to bring disease from above an epidemic threshold value to below the threshold value, thereby eliminating the threat of a large-scale epidemic. This can be achieved through interventions that either directly impact the infectiousness of the pathogen, modify patterns of interaction so that the pathogen cannot easily spread through the population, or immunize segments of the population. We call these three forms of intervention *transmission reducing*, *contact reducing*, and *immunizing* [PMS05].

Transmission reducing interventions introduce physical barriers to interrupt the spread of respiratory droplets or other infectious particles (e.g. face masks, gowns and gloves, hand hygiene, disinfection of animate objects). These interventions are modeled by reducing Tij, the probability of transmission from vertex i to vertex j, at an appropriate subset of vertices.

Contact reducing interventions include isolation of infected persons, quarantine of persons during their incubation period, patient and/or staff cohorting in hospitals, and closing public spaces (e.g. schools). These interventions are modeled by removing appropriate edges between vertices. For example, one can model school closures in an urban contact network by removing all edges that represent contacts between students, teachers, staff, etc., that take place during school.

Immunizing interventions include prophylactic medication and diverse vaccination strategies (e.g. ring vaccination, vaccinating individuals in contact with the identified infected case; targeted vaccination, vaccinating specific groups of individuals based on risk factors such as age, health, and place of employment; and general vaccination). Vaccination prior to an outbreak is tantamount to removing the immunized individuals from the network entirely and thus is modeled by removing vertices corresponding to those individuals.

To evaluate candidate control measures, we first modify the contact network accordingly and then quantify the impact of these changes on the size of an outbreak and demographic distribution of infections, identify segments of the population where compliance is most critical to successful control, and predict the individual and social benefits of complying with control measures. We will conclude with two practical examples.

Example 1: Controlling walking pneumonia outbreaks in closed settings.
Walking pneumonia is a relatively mild form of pneumonia that spreads rapidly in closed settings such as hospitals, nursing homes, military communities, and college campuses. As with many diseases, conducting human experiments to test control measures is often infeasible or unethical. In collaboration with U.S. Centers for Disease Control and Prevention (CDC) officials, we built some of the first network-based models of health-care settings with which we evaluated candidate

strategies [MNMS03]. This work led the CDC to reject prior candidate strategies (including isolation of sick patients and antibiotic prophylaxis) in favor of the following intuitive yet previously overlooked strategy: upon the first diagnosis of walking pneumonia, reassign caregivers so that they limit their interactions to fewer wards. Although intuitive in retrospect, this insight came directly from analyzing disease transmission in a realistic model of the underlying network of interactions.

Example 2: Optimal distribution of influenza vaccines. Pandemic influenza is characterized by wide geographic person-to-person spread of a novel strain toward which the population has no immunity. The three major pandemics of the 20th century—in 1918, 1957 and 1968—collectively caused at least 600,000 deaths in the U.S. and over 40 million deaths worldwide. Between major pandemics, the U.S. experiences seasonal outbreaks of interpandemic flu that kill over 30,000 people annually.

The threat of another avian influenza pandemic and the 2004-2005 influenza vaccine supply shortage in the United States has sparked a debate about optimal vaccination strategies to reduce the burden of morbidity and mortality caused by the influenza virus. During the 2004-2005 vaccine shortage, priority was given to individuals most at risk for hospitalization and death: healthy infants, the elderly, and individuals with chronic illnesses. These demographics, however, are not the primary spreaders of the influenza virus. Influenza outbreaks hinge, instead, on transmission by healthy schoolchildren [LKMF82, FHCF82, JM78, TPGC82], college students, and employed adults who are mobile and have many daily contacts [Gle96]. Some epidemiologists have therefore suggested that vaccines be given first to school-age children rather than to at-risk groups [LHNY04, LH05, WEH05].Vaccination of the children who are typically responsible for propagating disease throughout the community may slow the spread of disease and thereby indirectly decrease infection and mortality rates in the at-risk groups. This strategy has been supported both theoretically and empirically [MKL85, RSF01]. One modeling study suggests that vaccinating 80% of all schoolchildren is almost as effective as vaccinating 80% of the entire population [LHNY04].

Using an urban contact network based on demographic data for the city of Vancouver (with 260,000 individuals in 100,000 households), we quantitatively compared these two strategies for both interpandemic and pandemic influenza [BPMss]: (1) a mortality-based strategy that targets demographics with highest mortality rates (infants, elderly, and health-care workers for interpandemic flu; and infants, adults, and health-care workers for pandemic flu) and (2) a morbidity-based strategy that targets school-aged children [BPMss].

In contrast to prior studies [LHNY04], this study considers a relatively large population and the entire spectrum of viral transmission rates estimated for various influenza strains. As illustrated in Figure 5, the optimal strategy appears to depend critically on the viral transmissibility (reproductive rate) of the virus, with morbidity-based strategies outperforming mortality-based strategies for moderately transmissible strains, while the reverse is true for highly transmissible strains. This result holds for both interpandemic flu and pandemic flu. Furthermore, delays in vaccination and multiple introductions of disease into the community decrease the relative effectiveness of morbidity-based strategies. Thus, mortality-based strategies may be the prudent choice for outbreaks of new or atypical strains of influenza, when public health officials may not have reliable estimates for all (or any) of the
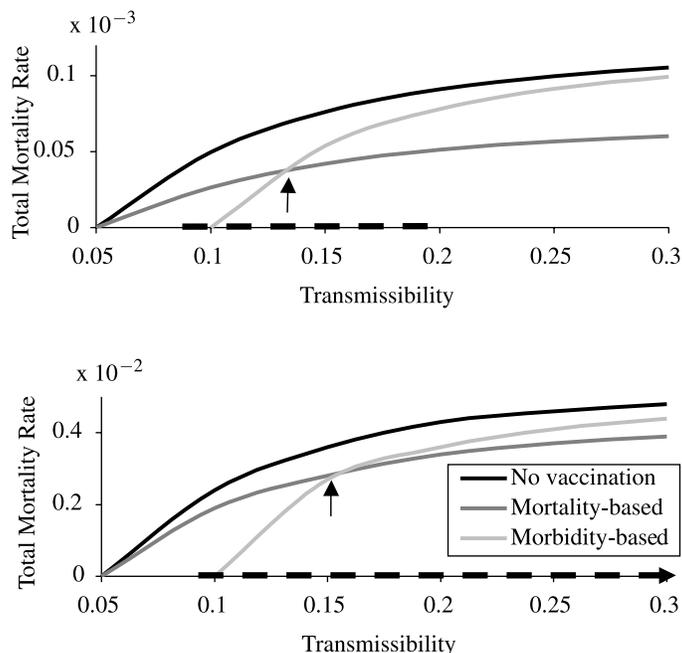
FIGURE 5. Expected mortality under different vaccination strategies for epidemic (top) and pandemic (bottom) influenza. Arrows mark transition between mildly contagious diseases which are better controlled by morbidity-based interventions and more highly contagious diseases which are better controlled by mortality-based interventions. The dashed regions along the x-axis give the range of transmission rates estimated from influenza data from annual epidemics (top) and the 1918 pandemic (bottom).

first three inputs and vaccination may be delayed. When reliable estimates of the key inputs are available significantly prior to an outbreak, then this approach can be applied to design optimal (rather than just prudent) priorities.

CONCLUSIONS

Mathematical epidemiology continues to evolve, offering more detailed forecasting and more effective control. Much of the recent progress has been fueled by the importation of relatively simple ideas from dynamical systems, probability theory and statistical mechanics. Despite these steps forward, infectious disease control is more often than not based on intuition rather than quantitative reasoning. This is particularly true for newly emerging diseases, for which we know little about the natural history and epidemiology of the pathogen. The variable public health response to SARS provides a compelling example of such uncertainty. When SARS emerged as a global threat in March 2003 [DGL03], the WHO and other agencies issued travel warnings for affected cities. Hong Kong, Singapore and China closed schools [Bei03]. A U.S. university denied attendance to students from China [CNN03a]. Public health authorities worldwide put thousands under

quarantine [CDC03]. Patients were strictly isolated [CDC04], and specific hospitals in some cities were designated to receive SARS cases [CNN03b].

Contact network epidemiology can provide quantitatively grounded guidance for public health officials facing newly emerging diseases like SARS and avian influenza. Consider the previously cited analysis of flu vaccine strategies [BPMss]. In contrast to other published mathematical approaches to this problem, our analytical methods have two advantages: they simultaneously capture realistic diversity in contact patterns ignored by many compartmental models and shortcut the extensive computer simulations required by agent-based models. In general, these advantages enable highly detailed and systematic consideration across several disease strains and intervention strategies. In the case of flu, such a study yielded important insight. If faced with a limited vaccine supply for either an interpandemic or pandemic strain of flu, morbidity-based strategies (e.g., targeting schoolchildren) are predicted to outperform mortality-based strategies (e.g., targeting elderly and infants) for strains that are mildly contagious, while the reverse is true for moderately to highly contagious strains. Furthermore, mortality-based strategies are generally advisable for populations experiencing repeated introductions of disease from other communities or delayed vaccination. This suggests that the U.S. Centers for Disease Control's 2004 decision to prioritize the very young, the old, and the immunocompromised— those most at risk for complications from flu—is generally more prudent than the recently promoted alternative strategy of vaccinating schoolchildren.

This methodology also sheds light on the incompatability between early estimates of R_0 for SARS and the case count in China (discussed above). This likely stemmed from the anomalously high contact rates in the hospital and apartment building upon which the R_0 estimates were based. Equation (22) clarifies that the basic reproductive rate of a disease is context dependent; that is, it fundamentally depends on the contact patterns of the population through which it spreads. Thus, while the SARS estimates may be valid for unusually crowded settings, they probably do not hold for typical rural or urban communities in general, like those through which SARS initially spread in China.

This example suggests that the emphasis on estimating the R_0 for an infectious disease may be misguided. Estimating the average transmissibility T instead of R_0 may be more valuable. This means reporting not just the number of new infections per case, but also the total estimated number of contacts during the infectious period of that case. Given the primary role of contact tracing in infectious disease control, the relevant data is often collected. Unlike R_0, T can be justifiably extrapolated from one location to another even if the contact patterns are quite disparate. We offer a simple example to illustrate the benefits of measuring T. Suppose we measure $R_0 = 2.7$ in a hospital where the average individual comes in close contact with 100 other individuals. Then the probability that an individual will catch the disease from an infected contact is just 2.7% or, in network terms, $T = .027$. Now suppose the typical individual in the general population has 10 close contacts that could potentially lead to the spread of a disease. If we extrapolate $R_0 = 2.7$ to the general public, then we predict that, on average, 2.7 out of every 10 contacts or 27% of contacts become infected. However, if we extrapolate $T = .027$ to the general public, we still have only 2.7% of contacts becoming infected, which gives us a much reduced expectation for the spread of the disease.

In closing, mathematical epidemiology is a rapidly developing field that thrives on collaborations among scientists, mathematicians, and public health officials. Contact network epidemiology is a particularly promising approach in which progress is fueled by both scientific curiosity and public health concerns. As demonstrated by the variable response to SARS, there is need for greater quantitative reasoning in public health. The onus is on the modelers not only to make technical advances but also to demonstrate the utility and accessibility of our models.

About the author

Lauren Ancel Meyers is an assistant professor in the Section of Integrative Biology at the University of Texas at Austin. She works in mathematical and experimental population biology and was named as one of the top 100 global innovators under age 35 by the MIT *Technology Review*.

References

[Abb52]   H. Abbey, *An examination of the reed-frost theory of epidemics*, Human Biology **24** (1952), no. 3, 201–233.

[AM91]   R. M. Anderson and R. M. May, *Infectious Diseases of Humans, Dynamics and Control*, Oxford University Press, Oxford, 1991.

[AO04]   L. A. N. Amaral and J. M. Ottino, *Complex networks - augmenting the framework for the study of complex systems*, Eur. Phys. J. B. **38** (2004), 147–162.

[BA99]   A. L. Barabasi and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), no. 5439, 509. MR2091634

[Bai75]   N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, Hafner Press, New York, 1975. MR0452809 (56:11084)

[BB04]   D. Bernoulli and S. Blower, *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it*, Reviews in Medical Virology **14** (2004), no. 5, 275–288.

[Bec77]   N. Becker, *Estimation for discrete time branching processes with application to epidemics*, Biometrics **33** (1977), no. 3, 515–522. MR0654337 (58:31657)

[Bei03]   *Designated hospitals in Beijing meeting SARS treatment demands*, People's Daily Online, 2003.

[BMST97] F. Ball, D. Mollison, and G. Scalia-Tomba, *Epidemics with two levels of mixing*, The Annals of Applied Probability **7** (1997), 46. MR1428749 (99a:92033)

[BMT03]   C. M. Booth, L. M. Matukas, G. A. Tomlinson, A. R. Rachlis, D. B. Rose, H. A. Dwosh, S. L. Walmsley, T. Mazzulli, M. Avendano, P. Derkach, I. E. Ephtimios, I. Kitai, B. D. Mederski, S. B. Shadowitz, W. L. Gold, L. A. Hawryluck, E. Rea, J. S. Chenkin, D. W. Cescon, S. M. Poutanen, and A. S. Detsky, *Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area*, JAMA (2003), 289.21.JOC30885.

[BPMss]   S. Bansal, B. Pourbohloul, and L. A. Meyers, *A comparative analysis of influenza vaccination programs*, PLoS Medicine **3** (2006), e387.

[BRO02]   T. Britton and P. D. O'Neill, *Bayesian inference for stochastic epidemics in populations with random social structure*, Scandanavian Journal of Statistics **29** (2002), no. 3, 375–390. MR1925565

[CDC03]   CDC, *Efficiency of quarantine during an epidemic of severe acute respiratory syndrome–Beijing, China, 2003*, Morbidity and Mortality Weekly Report **52** (2003), no. 43, 1037–1040.

[CDC04]   CDC informational bulletin, *Public health guidance for community-level preparedness and response to severe acute respiratory syndrome (SARS) version 2, supplement D: Community containment measures, including non-hospital isolation and quarantine*, 2004. `http://www.cdc.gov/ncidod/sars/guidance/D/pdf/lessons.pdf`

[CHECC03]   G. Chowell, J. M. Hyman, S. Eubank, and C. Castillo-Chavez, *Scaling laws for the movement of people between locations in a large city*, Physical Review E **68** (2003), 066102.

[CNN03a]   CNN.com, *Berkeley turns away students from SARS-hit regions*, 2003.

[CNN03b]   CNN.com, *SARS closing Beijing schools*, 2003.

[DdJM98]   O. Diekmann, M. C. M. de Jong, and J. A. J. Metz, *A deterministic epidemic model taking account of repeated contacts between the same individuals*, J. Appl. Prob. **35** (1998), 448–462. MR1641833 (99k:92041)

[DGL03]   C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L-M. Ho, T-Q. Thach, P. Chau, K-P. Chan, T-H. Lam, L-Y. Tse, T. Tsang, S-H. Liu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong*, The Lancet (2003), 1.

[DH02]   K. Dietz and J. A. P. Heesterbeek, *Daniel Bernoulli's epidemiological model revisited*, Mathematical Biosciences **180** (2002), 1–21. MR1950745 (2003m:92070)

[Dur99]   R. Durrett, *Stochastic spatial models*, SIAM Review **41** (1999), no. 4, 677–718. MR1722998 (2000h:60086)

[EGK04]   S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, *Modelling disease outbreaks in realistic urban social networks*, Nature **429** (2004), 180–184.

[FBMBss]   M. Ferrari, S. Bansal, L. A. Meyers, and O. N. Bjornstad, *Network frailty and the geometry of herd immunity*, Proceedings of the Royal Society (London) B (in press).

[FG00]   N. M. Ferguson and G. P. Garnett, *More realistic models of sexually transmitted disease transmission dynamics: sexual partnership networks, pair models, and moment closure*, Sex. Transm. Dis. **27** (2000), no. 10, 600.

[FHCF82]   J. P. Fox, C. E. Hall, M. K. Cooney, and H. M. Foy, *Influenza virus infections in Seattle families, 1975-1979*, American Journal of Epidmiology **116** (1982), 212–227.

[FKG03]   C. P. Farrington, M. N. Kanaan, and N. J. Gay, *Branching process models for surveillance of infectious diseases controlled by mass vaccination*, Biostatistics **4** (2003), no. 2, 279–295.

[Gle96]   W. P. Glezen, *Emerging infections: pandemic influenza*, Epidemiology Reviews **18** (1996), no. 1, 64–76.

[Gra83]   P. Grassberger, *Critical behavior of the general epidemic process and dynamical percolation*, Mathematical Biosciences **63** (1983), no. 2, 157–172.

[Ham06]   W. H. Hamer, *Epidemic disease in England—the evidence of variability and persistency of type*, The Lancet **i** (1906), 733–739.

[Het00]   H. W. Hethcote, *Mathematics of infectious diseases*, SIAM Review **42** (2000), 599. MR1814049 (2002c:92034)

[HY84]   H. W. Hethcote and J. A. Yorke, *Gonorrhea Transmission Dynamics and Control*, Lecture Notes in Biomathematics, vol. 56, Springer-Verlag, Berlin, 1984. MR0766910 (86a:92002)

[JM78]   L. C. Jennings and J. A. R. Miles, *A study of acute respiratory disease in the community of Port Chalmers*, Journal of Hygiene **81** (1978), 67–75.

[KG99]   A. Kleczkowski and B. T. Grenfell, *Mean field-type equations for spread of epidemics: the 'small world' model*, Physica A **274** (1999), no. 1-2, 1–385.

[KM27]   W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics*, Proceedings of the Royal Society (London) A **115** (1927), 700–721.

[KRM97]   M. J. Keeling, D. A. Rand, and A. J. Morris, *Correlation models for childhood diseases*, Proceedings of the Royal Society (London) B **264** (1997), 1149–1156.

[KWM03]   M. J. Keeling, M. E. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell, *Modelling vaccination strategies against foot-and-mouth disease*, Nature **421** (2003), no. 6919, 136–42.

[LCC03]   M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray, *Transmission dynamics and control of severe acute respiratory syndrome*, Science (2003), 1086616.

[LCH03]   Y. S. Leo, M. Chen, B. H. Heng, C. C. Lee, N. Paton, B. Ang, P. Choo, S. W. Lim, A. E. Ling, M. L. Ling, B. K. Tay, P. A. Tambyah, Y. T. Lim, G. Gopalakrishna, S. Ma, L. James, P. L. Ooi, S. Lim, K. T. Goh, Sk. K. Chew, and C. C. Tan, *Severe acute respiratory syndrome - Singapore, 2003*, Morbidity and Mortality Weekly Report **52** (2003), no. 18, 405.

[LEA01]   F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, *The web of human sexual contacts*, Nature **411** (2001), 907–908.

[LEA03]   F. Liljeros, C. R. Edling, and L. A. N. Amaral, *Sexual networks: implications for the transmission of sexually transmitted diseases*, Microbes (2003).

[LH05]   I. M. Longini and M. E. Halloran, *Strategy for distribution of influenza vaccine to high-risk groups and children*, American Journal of Epidemiology **161** (2005), 303–306.

[LHNY04]   I. M. Longini, M. E. Halloran, A. Nizam, and Y. Yang, *Containing pandemic influenza with antiviral agents*, American Journal of Epidemiology **159** (2004), 623–633.

[LKMF82]   I. M. Longini, J. S. Koopman, A. S. Monto, and J. P. Fox, *Estimating household and community transmission parameters of influenza*, American Journal of Epidemiology **115** (1982), 736–751.

[LM01]   A. L. Lloyd and R. M. May, *Epidemiology. How viruses spread among computers and people*, Science **292** (2001), no. 5520, 1316.

[Lon88]   I. M. Longini, *A mathematical model for predicting the geographic spread of new infectious agents*, Math. Biosci. **90** (1988), 367. MR0958149 (89h:92052)

[LP89]   C. Lefevre and P. Picard, *On the formulation of discrete-time epidemic models*, Math. Biosci. **95** (1989), no. 1, 27–35. MR1001289 (90e:92068)

[MKL85]   A. S. Monto, J. S. Koopman, and I. M. Longini, *The Tecumseh study of illness. XIII. Influenza infection and disease, 1976-1981*, American Journal of Epidemiology **121** (1985), 811–822.

[MNMS03]   L. A. Meyers, M. E. J. Newman, M. Martin, and S. Schrag, *Applying network theory to epidemics: Control measures for mycoplasma pneumoniae outbreaks*, Emerging Infectious Diseases **9** (2003), no. 2, 204.

[MNP06]   L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, *Predicting epidemics on directed contact networks*, Journal of Theoretical Biology **240** (2006), 400–418.

[Mor95]   M. Morris, *Data driven network models for the spread of disease*, Epidemic Models: Their Structure and Relation to Data (D. Mollison, ed.), Cambridge University Press, Cambridge, 1995, pp. 302–322.

[MPN05]   L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham, *Network theory and SARS: predicting outbreak diversity*, Journal of Theoretical Biology **232** (2005), 71–81. MR2106112

[New02]   M. E. J. Newman, *Spread of epidemic disease on networks*, Physical Review E **66** (2002), no. 1, art. no. 016128. MR1919737 (2003e:60223)

[New05]   _____, *Threshold effects for two pathogens spreading on a network*, Physical Review Letters **95** (2005).

[Org03]   World Health Organization, *Severe acute respiratory syndrome (SARS)*, 2003.

[PMS05]   B. Pourbohloul, L. A. Meyers, D. M. Skowronski, M. Krajden, D. M. Patrick, and R. C. Brunham, *Modeling control strategies of respiratory pathogens*, Emerg. Infect. Dis. **11** (2005), no. 8, 1249–1256.

[PSV01]   R. Pastor-Satorras and A. Vespignani, *Epidemic spreading in scale-free networks*, Phys. Rev. Lett. **86** (2001), no. 14, 3200.

[PVJ98]   N. Paneth and P. Vinten-Johansen, *A rivalry of foulness: Official and unofficial investigations of the London cholera epidemic of 1854*, American Journal of Public Health **88** (1998), no. 10, 1545–1553.

[RFD03]   S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L-M. Ho, T-H. Lam, T. Q. Thach, P. Chau, K-P. Chan, S-V. Lo, P-Y. Leung, T. Tsang, W. Ho, K-H. Lee, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, *Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions*, Science (2003), 1086478.

[Rot01]   R. B. Rothenberg, *How a net works: implications of network structure for the persistence and control of sexually transmitted diseases and HIV*, Sexually Transmitted Diseases **28** (2001), 63–68.

[RSF01]   T. A. Reichart, N. Sugaya, D. S. Fedson, W. P. Glezen, L. Simonsen, and M. Tashiro, *The Japanese experience with vaccinating school-children against influenza*, New England Journal of Medicine **344** (2001), 889–896.

[RST97]   R. B. Rothenberg, C. Sterk, K. Toomey, J. Potterat, D. Johnson, M. Schrader, and S. Hatch, *Using social network and ethnographic tools to evaluate syphilis transmission*, Sexually Transmitted Diseases **25** (1997), no. 3, 154–160.

[Sma02]   *Smallpox: A great and terrible scourge*, 2002, National Library of Medicine (National Institutes of Health) website: `http://www.nlm.nih.gov/exhibition/smallpox/sp_variolation.html`

[SS88]   L. Sattenspiel and C. P. Simon, *The spread and persistence of infectious diseases in structured populations*, Math. Biosci. **90** (1988), 341. MR0958148 (89m:92045)

[SWS02]   L. M. Sander, C. P. Warren, I. M. Sokolov, C. Simon, and J. Koopman, *Percolation on heterogeneous networks as a model for epidemics*, Mathematical Biosciences **180** (2002), 293–305. MR1950759 (2003m:92080)

[TPGC82]   L. H. Taber, A. Paredes, W. P. Glezen, and R. B. Couch, *Infection with influenza A/Victoria virus in Houston families, 1976*, Journal of Hygiene **86** (1982), 303–313.

[VdPvVdV98]   C. P. B. Van der Ploeg, C. van Vliet, S. J. de Vlas, J. O. Ndinya-Achola, L. Fransen, G. J. van Oortmarssen, and J. D. F. Habbema, *A microsimulation model for decision support in STD control*, Interfaces **28** (1998), 84–100.

[Volss]   E. Volz, *SIR dynamics in structured populations with heterogeneous connectivity*, Journal of Mathematical Biology (in press).

[Wat99]   D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, 1999. MR1716136 (2001a:91064)

[WEH05]   D. Weycker, J. Edelsberg, M. E. Halloran, I. M. Longini, A. Nizam, V. Ciuryla, and G. Oster, *Population-wide benefits of routine vaccination of children against influenza*, Vaccine **23** (2005), no. 10, 1284–1293.

[XHE04]   R-H. Xu, J-F. He, M. R. Evans, G-W. Peng, H. E. Field, D-W. Yu, C-K. Lee, H-M. Luo, W-S. Lin, P. Lin, L-H. Li, W-J. Liang, J-Y. Lin, and A. Schnur, *Epidemiologic clues to SARS origin in China*, Emerg. Infect. Dis. **10** (2004), no. 6.

Section of Integrative Biology, and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas 78712
    E-mail address: `laurenmeyers@mail.utexas.edu`