

Network theory and SARS:

Predicting outbreak diversity

Lauren Ancel Meyers^{1,2,†,*}, Babak Pourbohloul^{3,*}, M. E. J. Newman^{2,4},

Danuta M. Skowronski³, and Robert C. Brunham³

1. Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712
2. Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
3. University of British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, British Columbia, V5Z 4R4
4. Center for the Study of Complex Systems, University of Michigan, Randall Laboratory, 500 E. University Ave., Ann Arbor, MI 48109-1120

* These authors contributed equally to the conception and delivery of this work.

† Corresponding author: ancel@mail.utexas.edu, Ph: (512) 471-4950, Fax: (512) 471-3878

Abstract

Many infectious diseases spread through populations via the networks formed by physical contacts among individuals. The patterns of these contacts tend to be highly heterogeneous. Traditional “compartmental” modeling in epidemiology, however, assumes that population groups are fully-mixed, that is, every individual has an equal chance of spreading the disease to every other. Recent applications of compartmental models to Severe Acute Respiratory Syndrome (SARS) have resulted in estimates of the fundamental quantity called the basic reproductive number R_0 – the number of new cases of SARS resulting from a single initial case – to be above one, implying that, without public health intervention, all outbreaks should spark large-scale epidemics. Here we explain the inconsistency between these predictions and the observed epidemiology. We apply the powerful quantitative methods of network epidemiology to illustrate that for a single value of R_0 , any two outbreaks, even in the same setting, may have very different epidemiological outcomes. We offer quantitative insight into the heterogeneity of SARS outbreaks worldwide, and illustrate the utility of this approach for assessing public health strategies.

Introduction

Nearly one year since the first case of severe acute respiratory syndrome (SARS), a respiratory illness caused by a novel coronavirus, occurred in Guangdong province of China (November, 2002) and more than six months since the syndrome was first recognized outside of Asia (in Canada on March 13, 2003), its pattern of spread remains an enigma to public health officials and epidemiologists (1-6). Mathematical epidemiologists originally estimated the average number of secondary cases emanating from one primary case in a susceptible population (R_0) to be in the range of 2.2 and 3.6 for this virus – a high estimate, approximating that of a new subtype of influenza (7-9).

Despite this estimate and near-universal susceptibility, SARS has not emerged as a global pandemic. Instead, initial seeding was followed by intense but tightly circumscribed activity in some locales with only scant activity in others. In Canada, for instance, Toronto, Ontario and Vancouver, British Columbia were first affected nearly simultaneously in March 2003. By June 3, 2003, Toronto had experienced more than 209 probable cases and Vancouver had experienced only four probable cases. No other province of Canada reported any probable cases (10). The United States with a population more than fifty-fold greater than Toronto reported 69 probable cases – 67 imported and only two from secondary spread(11).

The discrepancy between the estimates of R_0 and the observed epidemiology might stem from early and effective intervention. Yet, even during the three and a half months of SARS spread in China between its initial appearance and the broad implementation of public health measures, case counts were much less than expected

from such values of R_0 . By definition, the total number of expected cases of a disease goes up by a factor of R_0 for every generation of infection, a generation being the mean time between an individual becoming infected and their infecting others. Based on recorded dates of the first symptoms for 124 pairs of subsequent infections in Singapore and Toronto, we estimate the average generation time for SARS to be 9.7 ± 0.3 days (12, 13). This estimate clearly depends on the accuracy the reported data. Roughly, the

cumulative number of SARS cases over D days should be $\sum_{i=0}^{D/\Delta} (R_0)^i = \frac{1 - R_0^{D/\Delta + 1}}{1 - R_0}$ (this is

capped by total population size and does not consider the reduction in R_t once a substantial proportion of the population is infected). Thus for R_0 ranging between 2.2 and 3.6, this equation predicts that in the first 120 days of transmission in China, there should have been approximately between 30,000 and 10 million. In fact only 782 cases were reported during the initial three months, which, using this simple calculation, suggests that R_0 should be much lower and closer to 1.6 (1). A subsequent estimate based on data from the Hong Kong and Singapore outbreaks brings R_0 down to 1.2, which, by the above formula, leads to an order of magnitude lower case counts (~ 50) than actually seen in the early stages of SARS in China (14).

Why do epidemiologists derive such varied estimates of R_0 and why were the initial estimates so high in comparison to the observed epidemiological patterns? We contend that the problem is two-fold. First, the basic premise of fully-mixed epidemiological models – that all individuals in a group are equally likely to become infected (or infect others) – often does not hold and therefore may lead to spurious

estimates or estimates that cannot justifiably be extrapolated from the specific setting in which they were measured to the broader community context. Early SARS estimates were based largely on transmission data from closed settings like hospitals and crowded apartment buildings, where there are unusually high rates of contact between individuals (7, 8). In fact, hospital transmission accounts for 50% of the value of R_0 described in (8). Contact rates in the population at large may be considerably lower and hence so may be the value of R_0 .

Secondly, even if these estimates are generally valid, $R_0 > 1$ will not guarantee that an outbreak will ignite an epidemic. This is contrary to the prediction of compartmental models where $R_0 > 1$ always predicts epidemic spread. Even compartmental models with greater complexity, where the population is stratified into many smaller subgroups based on infectiousness (8) or susceptibility (14), predict an epidemic with certainty for $R_0 > 1$. SARS, like many other infectious diseases exhibits great heterogeneity in transmission efficiency with certain individuals appearing to be responsible for large proportion of transmission events (12, 15, 16). These individuals may be “superspreaders” with unusually large numbers of contacts or “supershedders” who are unusually effective at excreting the virus into the environment they share with others. Traditional epidemiological theory assumes that each infected individual passes the disease on to exactly R_0 others. There is an enormous difference between this situation and one in which most infected individuals pass the disease on to only one or even zero others, but a small number pass it onto dozens or even hundreds – the mean value of R_0 can be the same in both cases, while the epidemiological outcomes are vastly different.

In this paper, we describe an analytical framework that captures the diverse interactions that underlie the spread of diseases such as SARS and thereby avoids the pitfalls of compartmental modeling. Given that public health control measures for communicable diseases, including contact tracing, isolation, quarantine, and ring vaccination (17, 18) have historically been predicated on sociologic considerations, models that share this perspective are long overdue. In fact, the basis of international SARS control has been founded upon the largely intuitive early interruption of critical social contacts (1).

Another alternative to compartmental modeling is "individual-based modeling", a primarily computational approach based on following the contact and infection histories of simulated individuals, and then aggregating the resulting data to make statistical predictions about disease outcomes. Most individual-based models assume pre-defined simple contact patterns such as regular lattices (19-22), although there have been recent efforts to consider more realistic contact patterns (23, 24). In contrast, our approach involves constructing networks based on information about real-life contacts between individuals and analyzing these networks to determine their crucial topological features. We then apply analytical tools to make epidemiological predictions and intervention recommendations based on these features. We refer to this approach, which makes no *a priori* assumptions about the global network structure and does not require extensive simulation, as contact network epidemiology (9, 25-33).

Contact Networks

Contact network models attempt to characterize every interpersonal contact that can potentially lead to disease transmission in a community. These contacts may take place within households, schools, workplaces, hospitals, etc. (Figure 1A). Each person in a community is represented as a *vertex* in the network and each contact between two people is represented as an *edge* (line) connecting their vertices. The number of edges emanating from a vertex, that is, the number of contacts a person has, is called the *degree* of the vertex. The distribution of the numbers of contacts – the degree distribution – is a fundamental quantity in network theory.

In the studies described here, we start by generating a plausible contact network for an urban setting using computer simulations. The simulations are based on data for the city of Vancouver, British Columbia. We choose $N=1000$ households at random from the Vancouver household size distribution (34), which yields approximately 2600 people. Household members are given ages according to the measured Vancouver age distribution (35), and, based on age, are then assigned to schools according to school and class size distributions (36), to occupations according to (un)employment data (37), to hospitals as patients and caregivers according to hospital employment and bed data (38), and to other public places. Within each location we create random connections between individuals with probabilities ranging from one for households to 0.3 for hospitals to 0.003 for other public places. Each school or hospital is sub-divided into classrooms or wards. Pairs of students or patients within these sub-units were connected with higher probability than pairs associated with different sub-units. Teachers are assigned to

classrooms and connected stochastically to appropriate students. Caregivers are assigned wards and then connected to appropriate patients. There are also low probability neighborhood contacts between individuals from different households.

This network offers a high degree of realism but is quite complex. We therefore use two simpler networks to provide additional insight. One is a random network with a Poisson degree distribution in which individuals connect to others independently and uniformly at random (Figure 1B) (39). Neither the simulated nor the Poisson network, however, contains a significant fraction of superspreaders. Therefore we also study a network with a power-law degree distribution (Figure 1C), a form much discussed in recent work on network epidemiology (40). This network has a "heavy tail" of superspreaders (Figure 2) and, as we will see, these individuals can have a profound effect on outbreak patterns despite being few in number. Network theorists often refer to such networks as scale-free because of the absence of a typical degree in the network.

We choose the parameters of the Poisson and power law networks so that all three networks share the same epidemic threshold T_c . This threshold, defined mathematically below, is the minimum disease transmission required for an outbreak to become a large-scale epidemic. Let p_k denote the probability that a randomly selected individual in a

network has degree k . Then, the Poisson network is given by $p_k = \frac{\bar{c}^k}{k!} \exp(-\bar{c})$ with

mean contact number $\bar{c}=19.6$; and the truncated power law network is given by

$$p_k = \begin{cases} 0 & \text{for } k = 0 \\ Ck^{-\beta} \exp(-\frac{k}{\gamma}) & \text{for } k > 0 \end{cases} \text{ with distribution parameters } \beta=94.2 \text{ and } \gamma=2. \text{ Truncation}$$

raises the epidemic threshold of the network to values comparable to those found for urban networks.

To generate the two idealized networks, we begin with a specified number of vertices and choose degrees for these vertices at random from the desired degree distribution. Then we connect random pairs of vertices, until the chosen degrees are exhausted. This often yields imperfect graphs with loops connecting vertices to themselves or redundant edges that connect the same two vertices more than once. We remove these imperfections using an algorithm suggested by Maslov et al. (41) in which we select at random two edges connecting, for example, vertex pairs AB and CD, and swap them so that they now connect AC and BD, unless this would create a new loop or double edge, in which case we do nothing. This process occasionally eliminates loops and repeated edges and by repeating it a sufficiently large number of times (depending on the network size) we can produce a network with none at all.

Epidemiological Analysis

Given the degree distribution of a network, we can use tools based on percolation theory (27, 40) to predict the fate of an outbreak of an infectious disease as a function of its transmissibility T . T is the average probability that an infectious individual will transmit the disease to a susceptible individual with whom they have contact. It summarizes core aspects of disease transmission including the rate at which contacts take place between individuals, the likelihood that a contact will lead to transmission, the duration of the

infectious period, and the susceptibility of individuals to SARS. It is related to the traditional R_0 according to

$$R_0 = T \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right)$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the mean degree and mean square degree of the network respectively. There is a critical transmissibility value

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

that corresponds to $R_0=1$, above which a population is vulnerable to large scale epidemics (but is not guaranteed to experience an epidemic) and below which only small local outbreaks occur.

Network theory makes a technical distinction between outbreaks and epidemics. An outbreak is a causally connected cluster of cases that, by chance or because the transmission probability is low, dies out before spreading to the population at large. In an epidemic, on the other hand, the infection escapes the initial group of cases into the community at large and results in population-wide incidence of the disease. The crucial difference is that the size of an outbreak is determined by the spontaneous dying out of the infection, whereas the size of an epidemic is limited only by the size of the population through which it spreads.

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

$$G_0(x) = \sum_{k=1} p_k x^k .$$

If we choose a random edge and follow it to the nearest vertex, then the pgf for the “excess degree” – the number of edges emanating from that vertex other than the one along which we arrived – is

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

The average degree and average excess degree equal the derivatives of these expressions

at $x=1$, that is, $\langle k \rangle = \sum_{k=1} k p_k$ and $\langle k_e \rangle = \frac{\sum_{k=1} k(k-1)p_k}{\sum_{k=1} k p_k} = \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right)$ respectively.

The value of the epidemic threshold T_c , the predicted average size of an outbreak $\langle s \rangle$ and probability of an epidemic S were first derived in (27). By nesting pgf’s for the number of new infections emanating from an infected vertex one can construct a pgf for the size of a small outbreak, and hence derive the average size of a small outbreak:

$$\langle s \rangle = 1 + \frac{T \langle k \rangle}{1 - T \langle k_e \rangle} .$$

This expression diverges when an outbreak becomes a large-scale epidemic. The epidemic threshold T_c (above) is the transmissibility value that marks this point.

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:

$$S = 1 - \sum_{k=1} p_k (1 + (u - 1)T)^k$$

where u is the probability that the person at the end of an edge does not have the disease and is solution to the equation

$$u = \frac{\sum_{k=1}^{\infty} k p_k (1 + (u-1)T)^{k-1}}{\sum_{k=1}^{\infty} k p_k}.$$

We use numerical root finding methods to solve for u . S is also the expected proportion of the population that will be infected should an epidemic occur.

Here we extend these analytics to predict the fate of an outbreak based on its initial conditions – how many and what kinds of individuals are already infected. We refer to the first individual in a community to come down with an infectious disease as *patient zero*. The probability that a patient zero with degree k will start an epidemic, π_k , is equal to the probability that transmission of the disease along at least one of the edges emanating from the original vertex will lead to an epidemic. For any one of its k edges, $1-T$ is the probability that the disease does not get transmitted along the edge and Tu is the probability that even if disease is transmitted to the next vertex, it does not proceed into a full-blown epidemic. Thus

$$\pi_k = 1 - (1 - T + Tu)^k.$$

The probability that an outbreak of size N will ignite an epidemic is $1 - \prod_{i=1}^N (1 - \pi_{k_i})$

where k_i is the 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 $(1 - \pi_{k_i})$'s replaced with the probability that a

typical infected individual does not spark an epidemic. The number of edges through which a typical infected individual can start an epidemic is given by the excess degree pgf, and the probability that one of those edges will not give rise to an epidemic is $1 - T + Tu$. Thus the probability that none of those edges will be a conduit to an epidemic

is $\sum_{k=1}^{\infty} k p_k (1 - T + Tu)^{k-1}$, and the probability that an outbreak of size N sparks an epidemic is

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

Finally, we derive an individual's risk of infection during an epidemic as a function of his or her degree. The probability \bar{q}_k that an individual with degree k will become infected during an epidemic is equal to one minus the probability that none of his or her k 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, $(1-u)(1-T)$, plus the probability that the contact was not infected in the first place, u . Thus, a randomly chosen vertex of degree k will become infected with probability

$$\bar{q}_k = \bar{q}_k = 1 - (1 - T + Tu)^k.$$

Epidemic simulations

We have simulated disease spread to verify our analytical predictions. Beginning with a susceptible population and a single case or a small cluster of cases, we iteratively take

each currently infected vertex, infect each of its susceptible neighbors with probability T and then change the status of the original vertex to recovered. This method of simulation does not capture the temporal progression of an epidemic, but just the overall number and distribution of infected individuals.

Predicting epidemics in the contact networks

Although the network and compartmental approaches may arrive at the same critical thresholds, they differ profoundly in their interpretations of these thresholds.

Compartmental models predict that any outbreak with $R_0 > 1$ will ignite an epidemic while network epidemiology predicts a specific probability S that an outbreak with $R_0 > 1$ will lead to an epidemic. S is often significantly less than one, and can be different for two networks with the same R_0 . When S is well below one and R_0 is well above one, it is very likely that communities with similar contact patterns will have diverse experiences with the disease, some experiencing large epidemics and other experiencing only minor outbreaks.

We illustrate this with the three networks described in Figures 1 and 2. We calculate S , the probability of an epidemic, as a function of T for all three networks (Figure 3). Above the epidemic threshold, the probability of an epidemic in the power law network and the probability of an epidemic in either of the other networks diverge quickly. Outbreaks are consistently less likely to reach epidemic proportions in the power law network than in the others. The vertical lines in Figure 3 correspond to estimates of R_0 for SARS (7, 8, 14). Whereas the compartmental approach predicts that all three

networks should experience epidemics at these values, we see that an epidemic is not guaranteed in any of them.

The different vulnerability of these networks arises from differences in the patterns of interpersonal contacts. For example, power law networks are made up primarily of vertices with very few contacts, and have a small minority of superspreaders of high degree. They will surpass the epidemic threshold when there is a non-negligible probability that an outbreak will reach members of this minority. Because superspreaders are rare, however, any particular outbreak may fail to reach this minority and thus never generate an epidemic. In contrast, vertices in a Poisson network will be fairly homogeneous and any two small outbreaks will be essentially equivalent. Whereas a power law network reaches an epidemic threshold when transmissibility is sufficiently high to reach superspreaders with reasonable frequency, a Poisson network only reaches an epidemic threshold when the ‘typical’ outbreak leads to an epidemic. Thus in the Poisson network, most outbreaks above the epidemic threshold give rise to epidemics.

We also predict the size of a small outbreak when $T < T_c$, an estimate that is not possible for compartmental models (Figure 3). Below the threshold, a typical outbreak in a power law network will die out after only one or two cases, whereas such outbreaks in the simulated urban or Poisson networks will include, on average, eight or ten individuals. Predicting the average size of a non-epidemic outbreak, $\langle s \rangle$, can aid in intervention against infectious diseases that rarely or never become self-sustaining epidemics but nevertheless have significant impact.

The similarity between the values of S and $\langle s \rangle$ predicted analytically and those measured through simulation (Figure 3) suggest that the analytical tools provide a powerful shortcut around computationally expensive epidemic simulations. While it may not be surprising that the randomly assembled power law and Poisson networks should fit the analytical predictions well, the simulated urban network introduces additional correlations in the network structure that are not considered in our analytic calculations.

The Impact of the Initial Conditions

Given that, for most diseases and most communities, the introduction of a single case may or may not spark an epidemic, we can ask what factors favor one or the other outcome (epidemic or no epidemic). Using the formulas described above, we predict the likelihood of an epidemic as a function of the degree of patient zero. Both simulation and analysis of the simulated urban network illustrate that the likelihood of an outbreak is a monotonically increasing function of the degree of patient zero (Figure 4A). Near the epidemic threshold, the risk increases exponentially with degree, whereas well above the epidemic threshold (e.g., $R_0=2.7$), the risk increases steeply until an epidemic is almost guaranteed. Therefore, even without differences in public health intervention, two identical communities can experience significantly different SARS outbreaks if the contact patterns of the first cases in each community differ.

In Figure 4B, we predict the probability of an epidemic based on the size of the initial outbreak. Intuitively, the more initial cases there are, the more precarious the outcome. For diseases far above the epidemic threshold, the threat of an epidemic is

overwhelming for even very small outbreaks. Thus vigilant tracking of case-contacts of the first few cases is paramount to the control of such a disease.

The Impact of Intervention

Along with the behavior of patient zero, individual and organized forms of disease intervention can dramatically affect an outbreak. Here we describe a simple application of our analytical toolkit to predicting the efficacy of control measures. There are two basic categories of intervention. *Transmission interventions*, like wearing face masks and washing hands, reduce the likelihood that a contact with another person leads to transmission of the disease. *Contact interventions*, like avoiding public places and rearranging the patterns of interaction between caregivers and patients in a hospital, eliminate or reduce the opportunity for such contacts.

Individuals can protect themselves by adopting such measures. If followed equally rigorously, an individual will benefit equally from the two forms of intervention. That is, by reducing one's contacts by a fraction a , an individual of degree k reduces his or her risk of infection from $1 - (1 - T + Tu)^k$ to $1 - (1 - T + Tu)^{ak}$; and reducing the likelihood of transmission per encounter by the same fraction a , reduces the risk to $1 - (1 - aT + aTu)^k$. To leading order, these two probabilities are equal. The efficacy of these strategies increases with the magnitude of the intervention a , decreases with the degree of the individual, and depends on the nature of the underlying contact network (Figure 5). Near the epidemic threshold, individuals of different degree benefit equally from such strategies. Well above the epidemic threshold, the more conservative one's

contact patterns, the greater the impact of such interventions. For very highly connected individuals, partial interventions will have little or no effect.

There are situations where one form of intervention is more feasible than the other. In such cases, instead of comparing transmission and contact reductions of equal compliance (as in Figure 5), one should compare transmission and contact interventions of equal feasibility. For example, the surgical masks seen on the streets of Hong Kong may be poorly effective against the SARS agent, and may hardly lower transmissibility, whereas the N-95 or surgical masks worn by hospital workers may be quite effective (42). Furthermore, individuals may easily reduce their social interactions by avoiding crowded shopping malls and movie theaters, whereas hospital caregivers may be unable to reduce the number of patients with whom they interact. Thus, taking into account intervention feasibility, members of the general public may be advised to avoid contact opportunities, whereas hospital employees should invest in high efficiency masks.

Policy makers take a more statistical approach to intervention than do individuals. Their goal is to reduce the likelihood and size of epidemics while minimizing cost to society. The most restrictive policies, such as closed borders and quarantining every member of society, are obviously economically catastrophic. Network theory allows us to predict quantitatively the epidemiological impacts of diverse strategies, which can then be paired with economic and sociological assessments of such strategies.

Widespread adoption of transmission interventions will lower T and may successfully bring a population under the epidemic threshold. For example, if a disease is spreading through our simulated urban network with $T=0.1550$ (equivalently, $R_0 = 2.7$),

then Figure 3 indicates that average transmissibility must be lowered by 62.9% to bring the population under the epidemic threshold (i.e., under $T=.0575$). An intervention that lowers transmission by less than 62.9% (for example, poor quality face masks) will not in itself prevent the emergence of an epidemic. We can also simply extend our analysis of the individual benefits of contact interventions to predict their effects on the probability and size of an epidemic (not shown here).

Discussion

Contact network modeling of SARS transmission allows us to step beyond the purview of the traditional compartment approach, and make quantitative predictions about the scale of outbreaks from information about the first few cases (Figure 4). This provides insight into the very different SARS outbreaks in Toronto and Vancouver. The first cases in both cities were exposed to SARS at Hotel M in Hong Kong on February 21, 2003. Patient zero in Toronto (onset February 25, 2003) was the matriarch of an extended, multi-generational family who died at home on March 5 as an unrecognized case of SARS (13). In addition to the five of ten persons within her family cluster who were affected, subsequent spread to health care workers, patients and their families has culminated in more than 200 probable cases in Toronto, most of them indigenously acquired (10, 13). In contrast, patient zero in Vancouver (onset February 26, 2003) returned from traveling with his wife to an otherwise empty abode on March 7, 2003 and was almost immediately hospitalized (13). Two of the three other probable cases in Vancouver were imported from abroad and no secondary case of SARS from patient zero was detected (10). Figure

4A illustrates that, in our simulated urban network, the difference between a patient zero with one to five contacts (Vancouver) and a patient zero with ten or more contacts (Toronto) can mean a doubling or tripling in the likelihood of an epidemic.

R_0 is a valuable epidemiological quantity. It is relatively straightforward to derive based on routinely collected epidemiologic data and can be predictive in homogeneous settings; it also has historically informed effective vaccination strategies (9, 33). Yet R_0 has its limits. Since R_0 is a function of both the transmissibility of a disease and the contact patterns that underlie transmission, then measuring R_0 in a location where contact rates are unusually high will lead to an estimate that is not appropriate for the larger community. As in the case of SARS, however, data is often only available for a limited and unrepresentative sample of the larger population.

Estimating the transmissibility T instead of R_0 gives us a way out of this difficulty. 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, whereas 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.

We have illustrated that percolation theory allows us to shortcut time-consuming simulations to produce robust predictions about the size and likelihood of an epidemic, the implications of patient zero's contact patterns and the initial size of an outbreak, and optimally reducing the risk of infection at a personal and population level. To truly reap these benefits of the network epidemiology, we must have not only good estimates of T but also realistic models of contact networks. We have presented a first step towards a realistic community network and have shown that it significantly departs from the idealized networks previously used in network epidemiology, and yet still is sufficiently random for the application of epidemiological network analysis. The task of making even more realistic contact network models of small communities, hospitals, cities, even global population is enormous, but promises a step change in our ability to predict and effectively control the spread of infectious disease. As we incorporate better data at these various scales, network theory will allow us to generalize our predictions and make better suggestions for epidemiological control.

Acknowledgements

The authors thank J.J. Bull, M. Girvan, and S. Meyers for useful conversations and editorial suggestions. This work was supported in part by grants from Canadian Institutes of Health Research (CIHR) to the BC SARS Investigators Collaborative that includes L.A.M, B.P., D.M.S., and R.C.B. and from the James S. McDonnell Foundation and the National Science Foundation (DMS-0234188) 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.

References

1. <http://www.who.int/csr/sars/en>.
2. J. S. M. Peiris *et al.*, *Lancet* **361**, 1319 (2003).
3. T. G. Ksiazek *et al.*, *N Engl J Med* **348**, 1953 (2003).
4. C. Drosten *et al.*, *N Engl J Med* **348**, 1867 (2003).
5. M. A. Marra *et al.*, *Science* **300**, 1399 (2003).
6. D. Cyranoski, A. Abbott, *Nature* **423**, 467 (2003).
7. M. Lipsitch *et al.*, *Science*, 1086616 (May 23, 2003, 2003).
8. S. Riley *et al.*, *Science*, 1086478 (May 23, 2003, 2003).
9. H. Hethcote, *SIAM Review* **42**, 599 (2000).
10. <http://www.sars.gc.ca>.
11. <http://www.cdc.gov/od/oc/media/sars.htm>
12. Y. S. Leo *et al.*, *Morbidity and Mortality Weekly Report* **52**, 405 (2003).
13. S. M. Poutanen *et al.*, *N Engl J Med* **348**, 1995 (May 15, 2003, 2003).
14. G. Chowell *et al.*, *Journal of Theoretical Biology* **224**, 1-8 (2003).
15. C. M. Booth *et al.*, *Jama*, 289.21.JOC30885 (May 6, 2003, 2003).
16. C. A. Donnelly *et al.*, *The Lancet*, 1 (May 7, 2003, 2003).
17. D. Greenhalgh, *Communications in Statistics: Stochastic Models* **2**, 339-363 (1986).
18. J. Müller *et al.*, *Journal of Mathematical Biology* **41**, 143-171 (2000).
19. R. Durrett, *SIAM Review* **41**, 677-718 (1999).
20. A. Kleczkowski, B. T. Grenfell, *Physica A* **274**, 1-385 (1999).

21. L. M. Sander *et al.*, *Mathematical Biosciences* **180**, 293-305 (2002).
22. T. Ritton, P. D. O'Neill, *Scandinavian Journal of Statistics* **29**, 375-390 (2002).
23. G. Chowell *et al.*, "Analysis of a real world network: The City of Portland" *Tech. Report No. BU-1604-M* (Department of Biological Statistics and Computational Biology, Cornell University, 2002).
24. C. P. B. Van der Ploeg *et al.*, *Interfaces* **28**, 84-100 (1998).
25. M. J. Keeling *et al.*, *Nature* **421**, 136-42 (Jan 9, 2003).
26. L. A. Meyers *et al.*, *Emerging Infectious Diseases* **9**, 204 (FEB, 2003).
27. M. E. J. Newman, *Physical Review E* **66**, art. no.-016128 (JUL, 2002).
28. L. Sattenspiel, C. P. Simon, *Math. Biosci.* **90**, 341 (1988).
29. F. Ball *et al.*, *The Annals of Applied Probability* **7**, 46 (1997).
30. M. Morris, in *Epidemic Models: their structure and relation to data* D. Mollison, Ed. (Cambridge University Press, Cambridge, 1995).
31. I. M. Longini, *Math. Biosci.* **90**, 367 (1988).
32. N. M. Ferguson, G. P. Garnett, *Sex Transm Dis* **27**, 600 (Nov, 2000).
33. A. L. Lloyd, R. M. May, *Science* **292**, 1316 (May 18, 2001).
34. Statistics Canada, <http://www.statcan.ca/english/Pgdb/famil53i.htm> (2001).
35. BC Stats, <http://www.bcstats.gov.bc.ca/data/dd/facsheet/mun.htm> (2001).
36. Vancouver School Board December 2002 Ready Reference, <http://www.vsb.bc.ca/board/publications.htm> (2002).
37. BC Stats, <http://www.bcstats.gov.bc.ca/data/lss/labour.htm> (2002).

38. *The British Columbia Health Atlas* (Centre for Health Services and Policy Research, Vancouver, B.C., Canada, ed. 1st, 2002).
39. A. L. Barabasi, R. Albert, *Science* **286**, 509 (Oct 15, 1999).
40. R. Pastor-Satorras, A. Vespignani, *Phys Rev Lett* **86**, 3200 (Apr 2, 2001).
41. S. Maslov, K. Sneppen, A. Zaliznyak. Preprint cond-mat/0205379 (2002).
42. W. H. Seto *et al.*, *The Lancet* **361**, 1519 (2003/5/3, 2003).

Legends

Figure 1. Schemata of (A) urban, (B) power law, and (C) Poisson networks. Dots represent individuals and lines between dots represent contacts between individuals that could potentially lead to disease transmission.

Figure 2. Cumulative degree distributions for simulated urban, Poisson, and power law networks. As described in the text, these share the same epidemic threshold (T_c). Each line gives the probability that a randomly chosen individual (vertex) will have at least the number of contacts (degree) indicated on the x-axis. The degree distribution for the urban network is nearly exponential for degrees greater than ten.

Figure 3. Predicting outbreaks and epidemics. The left graph illustrates the average number of people infected in a small outbreak, $\langle s \rangle$, when T is below the epidemic threshold. The right graph illustrates S , the probability that an epidemic occurs when T is above the epidemic threshold. S also equals the expected fraction of the population infected during an epidemic, should one occur. The vertical lines correspond to recent estimates for SARS (7, 8, 14). Simulation values are based on 2571 simulated epidemics, each starting with a unique individual in the network.

Figure 4. Predicting an epidemic from initial conditions. (A) The probability that patient zero will ignite a full-blown epidemic is monotonically increases with his or her degree. (B) The probability of a full-blown epidemic increases with the size of the initial

outbreak. This calculation does not assume knowledge of the specific degrees of the individuals affected in the outbreak, information that would improve the prediction. Both figures are based on the simulated urban network. Simulation values are based on 2571 simulated epidemics, one for every unique patient zero in the network. Discrepancy between simulations and analysis is likely caused by the finite size of the network, which contains very few high degree vertices, and by the intrinsic community structure in which high degree vertices (like teachers and caregivers) are preferentially connected to each other.

Figure 5. Individual intervention. The probability that an individual will become infected increases with the extent of personal precautions (calculated for a simulated urban network). For example, the 25% lines indicate the risk of infection if an individual who either reduces his/her contacts by 25% or lowers the likelihood of transmission per contact by 25%.

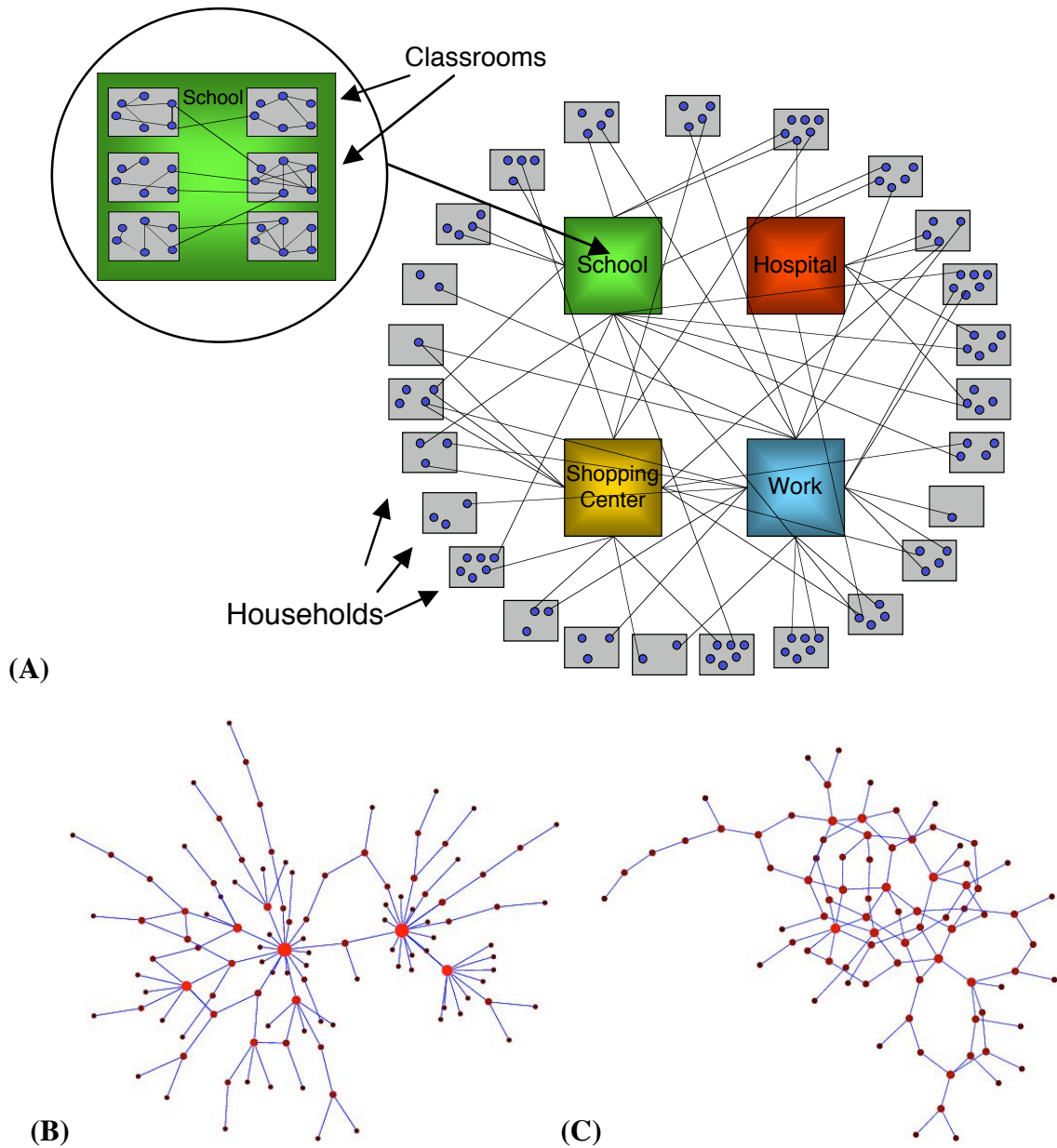


Figure 1

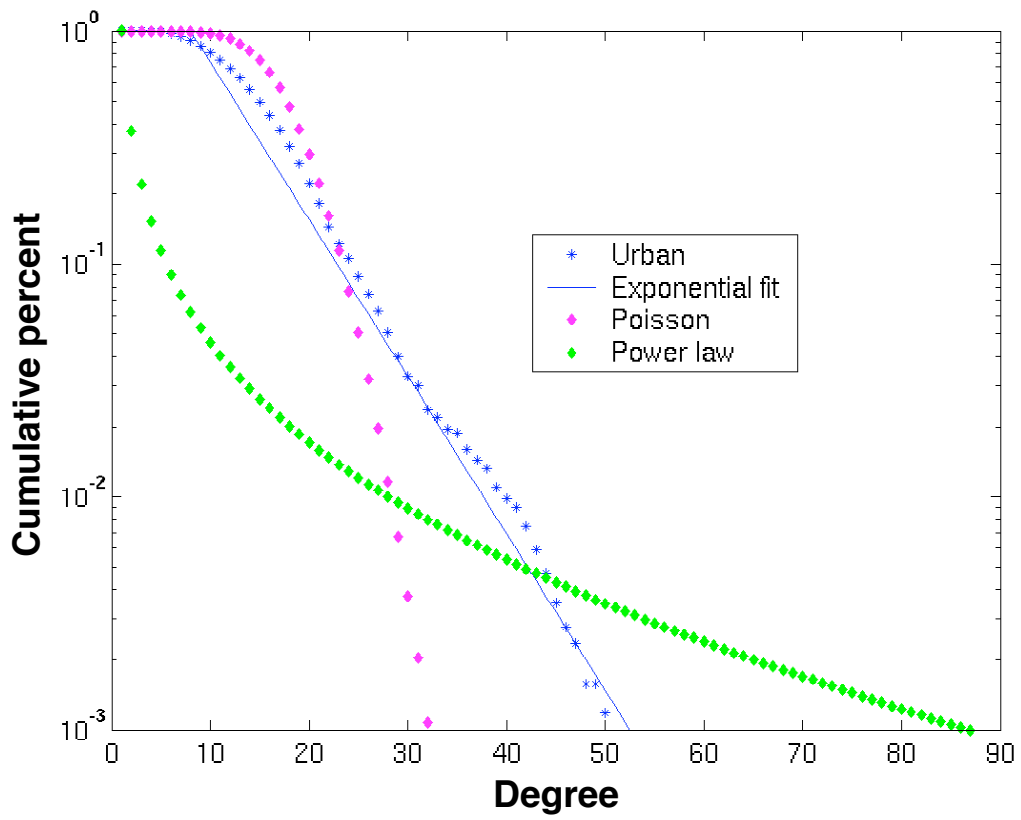


Figure 2

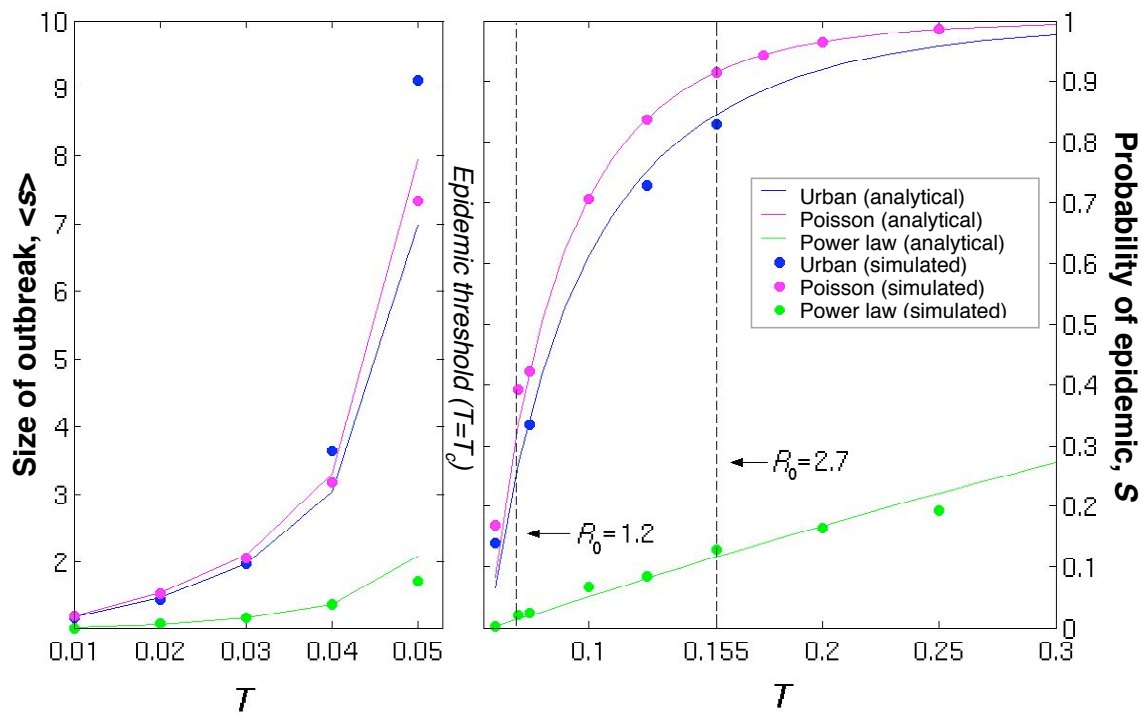
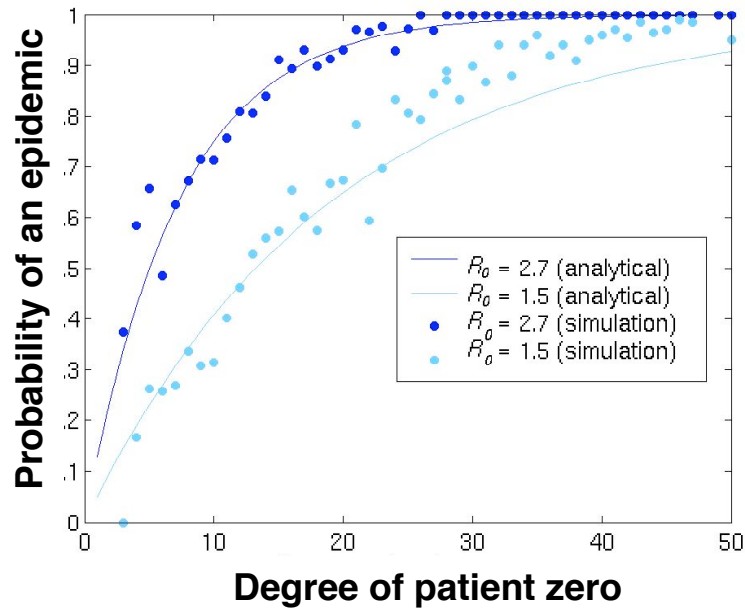
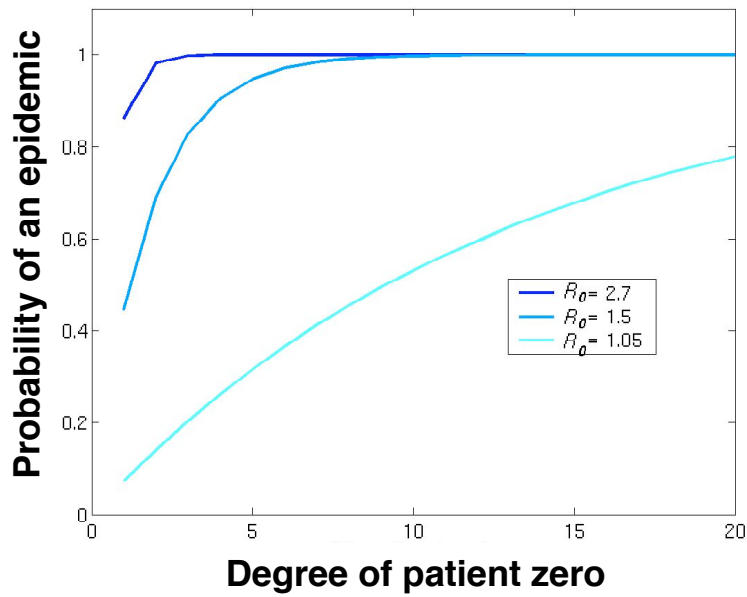


Figure 3



(A)



(B)

Figure 4

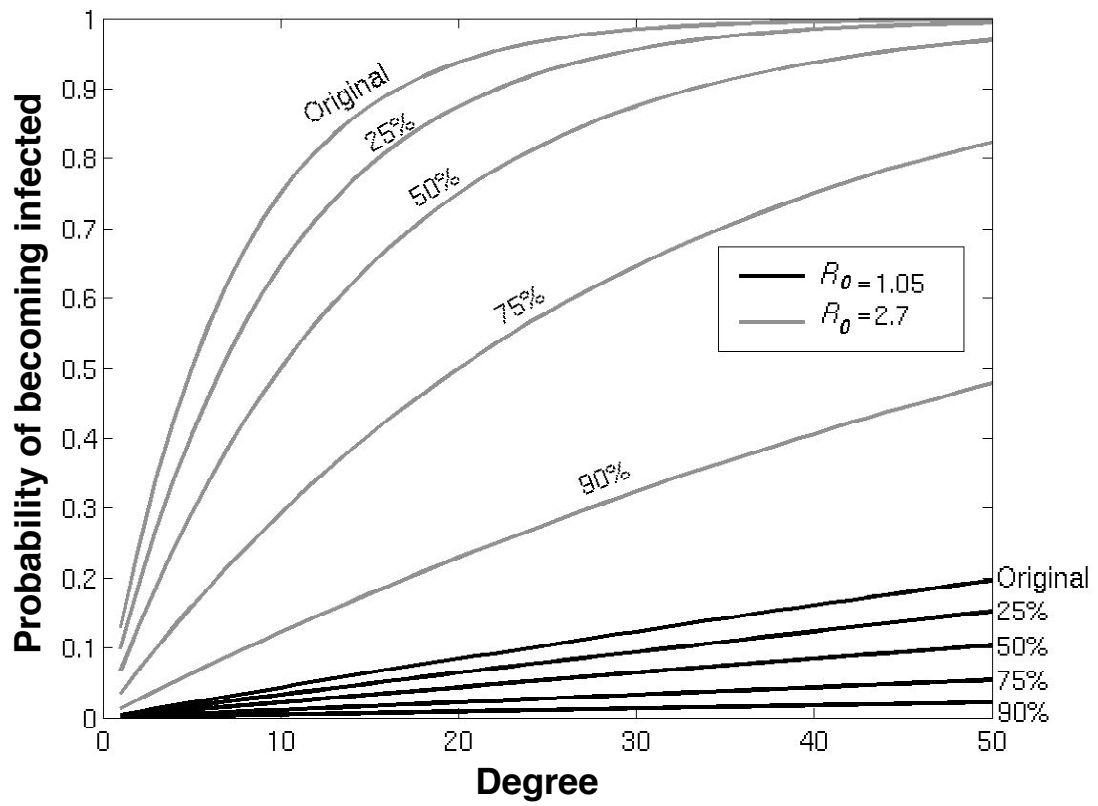


Figure 5