JOURNAL OF THE ROYAL SOCIETY Interface

J. R. Soc. Interface (2007) **4**, 879–891
doi:10.1098/rsif.2007.1100
Published online 19 July 2007

When individual behaviour matters: homogeneous and network models in epidemiology

Shweta Bansal[1], Bryan T. Grenfell[2,3] and Lauren Ancel Meyers[4,5,*]

[1]*Computational and Applied Mathematics, Institute for Computational Engineering and Sciences, University of Texas at Austin, 1 University Station, C0200, Austin, TX 78712, USA*
[2]*Center for Infectious Disease Dynamics, Biology Department, 208 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16802, USA*
[3]*Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA*
[4]*Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station, C0930, Austin, TX 78712, USA*
[5]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Heterogeneity in host contact patterns profoundly shapes population-level disease dynamics. Many epidemiological models make simplifying assumptions about the patterns of disease-causing interactions among hosts. In particular, homogeneous-mixing models assume that all hosts have identical rates of disease-causing contacts. In recent years, several network-based approaches have been developed to explicitly model heterogeneity in host contact patterns. Here, we use a network perspective to quantify the extent to which real populations depart from the homogeneous-mixing assumption, in terms of both the underlying network structure and the resulting epidemiological dynamics. We find that human contact patterns are indeed more heterogeneous than assumed by homogeneous-mixing models, but are not as variable as some have speculated. We then evaluate a variety of methodologies for incorporating contact heterogeneity, including network-based models and several modifications to the simple SIR compartmental model. We conclude that the homogeneous-mixing compartmental model is appropriate when host populations are nearly homogeneous, and can be modified effectively for a few classes of non-homogeneous networks. In general, however, network models are more intuitive and accurate for predicting disease spread through heterogeneous host populations.

Keywords: epidemic model; compartmental model; homogeneous-mixing; contact network

1. INTRODUCTION

Epidemics caused by the transmission of infectious agents are marked by variation. Heterogeneities in pathogens, host populations and the interactions between them profoundly affect the dynamics of infection. For example, in 1984, in a study of the spread of gonorrhoea in the United States, Hethcote and Yorke showed that 60% of all infections were caused by a small group of individuals making up only 2% of the entire population (Hethcote & Yorke 1984). In 1999, a cluster of five cases of measles in a small elementary school in The Netherlands led to an outbreak of approximately 3000 cases, despite the fact that 96% of the population of The Netherlands was vaccinated against measles at the time (Centers for Disease Control and Prevention 2000). In the spring of 2003, two individuals travelling from Vancouver and Toronto, respectively, were infected almost simultaneously with SARS by a source in Hong Kong. There were no secondary cases reported from the individual from Vancouver; in contrast, the infected individual from Toronto went on to infect five other people, which led to an outbreak of 200 cases in Toronto (Poutanen *et al.* 2003).

There are several epidemiologically important sources of variability, including disease-independent host parameters—age, sex, contact rate and compliance to public health recommendations—and disease-dependent host parameters—susceptibility to disease, transmission rate, mode of transmission and recovery rate. In this article, we will focus on disease-independent heterogeneity in host contact rates. That is, the number of potentially disease-causing interactions can vary widely across a host population. Although this is only one aspect of host

*Author and address for correspondence: Section of Integrative Biology, University of Texas at Austin, 1 University Station, C0930, Austin, TX 78712, USA (laurenmeyers@mail.utexas.edu).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2007.1100 or via http://www.journals.royalsoc.ac.uk.

One contribution of 20 to a Theme Issue 'Cross-scale influences on epidemiological dynamics: from genes to ecosystems'.

Received 11 May 2007
Accepted 22 June 2007

heterogeneity, it is an important one. Variability in contact patterns can stem from social structure, age, sex, spatial structure and behavioural differences. This heterogeneity is ubiquitous at many scales and can cause variability in disease parameters such as infectivity and susceptibility (Hethcote & Yorke 1984; Addy *et al.* 1991). Such individual-level diversity can profoundly shape population-level disease dynamics.

The simplest of the traditional mathematical models for the spread of infectious disease typically does not take such diversity into account, but assumes that communities are homogeneous; that is, they are made up of individuals who mix uniformly and randomly with each other. This simplifying assumption makes the analysis tractable but may not adequately reflect reality. Despite this assumption, these *homogeneous-mixing compartmental models* have proved to be robust and predictive (Anderson & May 1992; Mollison *et al.* 1993). Although we will focus on the simplest compartmental model, we note that this approach has been successfully extended to capture large-scale host heterogeneities. These extensions of the simple compartmental framework have included age-specific contact patterns and heterogeneities induced by spatial structure (Ball *et al.* 1997; Bjørnstad *et al.* 2002; Grenfell *et al.* 2002), but they do not allow for individual-level resolution.

After nearly a century of successes with these models, mathematical epidemiologists have turned their attention to individual-based approaches and specifically to network modelling. These new approaches, spurred by the availability of data and the maturation of network theory, reject the homogeneous-mixing assumption and explicitly capture the diverse patterns of interaction that underlie disease transmission (Barbour & Mollison 1990; Watts & Strogatz 1998; Pastor-Satorras & Vespignani 2001; Newman 2002; Meyers *et al.* 2005; Shirley & Rushton 2005).

Here, we reconcile the historical successes of homogeneous-mixing compartmental models with the advantages of new individual-based approaches. We start with a brief introduction to the contact network framework and then translate the homogeneous-mixing assumption into network terminology. Using several real and simulated datasets of human contact networks, we show that the homogeneous-mixing assumption may give rough but reasonable approximations for many populations. We then characterize the 'epidemiological distance' between these empirical networks and the homogeneous-mixing assumption of the simple model. Finally, we review and evaluate a variety of methodologies for incorporating contact heterogeneity, including several modifications to the simplest compartmental framework. We emphasize that many of the new network-based approaches are as mathematically tractable as the simplest compartmental models; although the compartmental framework can be extended to consider many complexities, the elegance with which the network-based approaches incorporate heterogeneity makes them an attractive option for future epidemiological studies.

2. AN INTRODUCTION TO CONTACT NETWORKS

Many diseases spread through human populations via close physical interactions. The interpersonal contact patterns that underlie disease transmission can naturally be thought to form a network, where links join individuals who interact with each other. During an outbreak, disease then spreads along these links. All epidemiological models make assumptions about the underlying network of interactions, often without explicitly stating them. Contact network models, however, mathematically formalize this intuitive concept so that epidemiological calculations can explicitly consider complex patterns of interactions.

Formally, a contact (or social) network model explicitly represents host interactions that mediate disease spread. A *node* in a contact network represents an individual host, and an *edge* between two nodes represents an interaction that may allow disease transmission. A node's *degree* is the number of edges attached to it (i.e. its number of contacts) and the *degree distribution* of a network is the frequency distribution of degrees throughout the entire population.

An exact contact network model requires knowledge of every individual in a population and every disease-causing contact between individuals (e.g. sneezing in the case of airborne diseases or sexual contact in the case of sexually transmitted diseases). For even small populations, this is typically unfeasible, and thus researchers typically work with approximate networks. There are several techniques for gathering the information needed to build realistic contact network models. They include tracing all infected individuals and their contacts during or following an outbreak (e.g. Klovdahl *et al.* 1977), surveying individuals in populations (e.g. Eubank *et al.* 2004) and using census (e.g. Meyers *et al.* 2005), social characteristic (e.g. Halloran *et al.* 2002) or other collected data (e.g. Meyers *et al.* 2003).

Characterizing network structure has become a multidisciplinary cottage industry, with researchers across epidemiology, sociology, biology, computer science and physics searching volumes of data for meaningful patterns. Researchers often look for global statistical properties in network data, and have paid special attention to small-world networks—characterized by high levels of both local clustering and global connectivity (Watts & Strogatz 1998)—and scale-free networks—characterized by degree distributions that follow a power-law distribution with a small fraction of very highly connected hubs (Barabási & Albert 1999). Scale-free networks have been reported in many technological (e.g. the Internet, the World Wide Web) and biological systems (e.g. metabolic, protein interactions, transcription regulation and protein domain; Albert *et al.* 1999; Faloutsos *et al.* 1999; Wuchty 2001; Giot *et al.* 2003; Bork *et al.* 2004; Hatzimanikatis *et al.* 2004; Luscombe *et al.* 2004). These highly structured networks are often contrasted with three classes of 'null' networks: (i) *lattices* in which all nodes have the same degree, and any given node is connected to physically proximate nodes, (ii) *regular random networks* in which all nodes have the same degree, but any given node is connected to

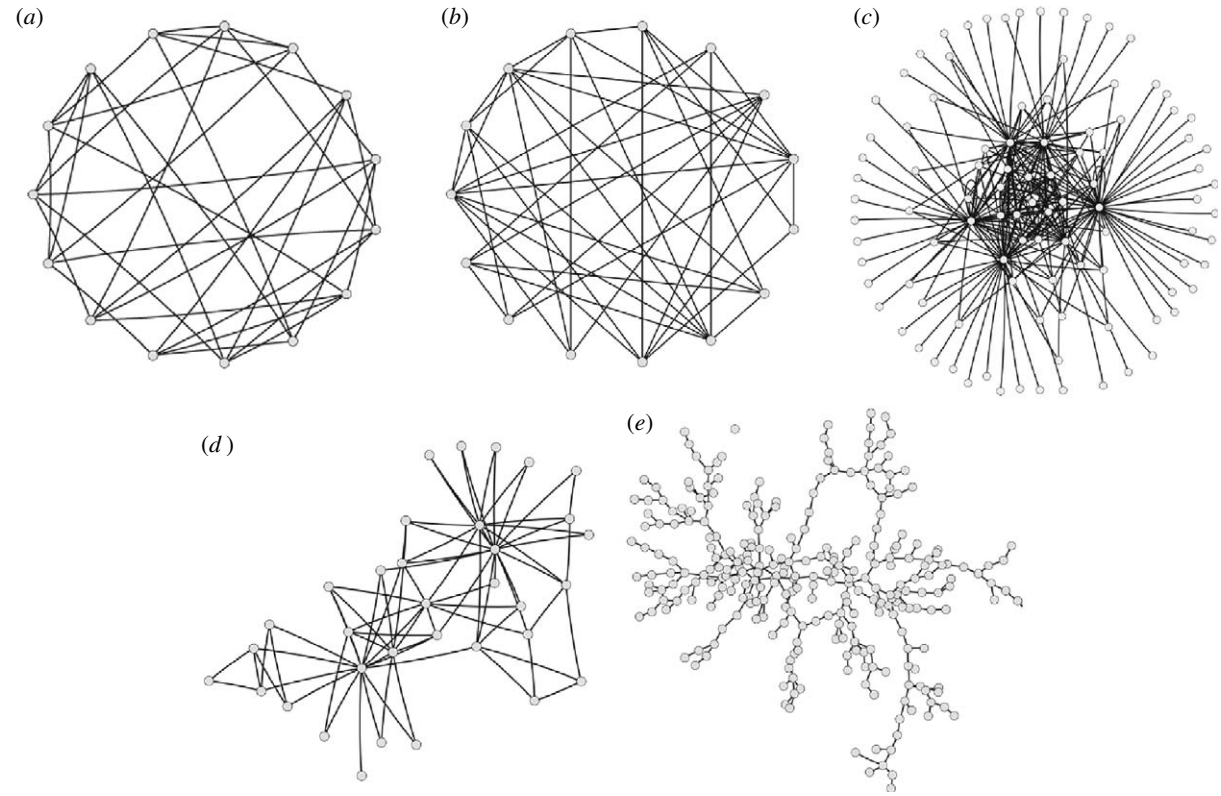

Figure 1. Examples of (*a*) a regular random network with 15 nodes and mean = 5, (*b*) a Poisson random graph with 15 nodes and mean = 5, (*c*) a scale-free random graph with 100 nodes and mean = 5, (*d*) the Zachary Karate Club contact network (Zachary 1977) with 34 nodes and mean ≈ 5 and (*e*) the sexual network for adolescents in a Midwestern US town, with 287 nodes and mean ≈ 2. (These networks do not contain spatial information, and the layouts were chosen simply to facilitate visual comparisons.)

randomly chosen nodes throughout the network, and (iii) *Poisson random networks* (also called Erdos–Renyi random graphs) in which some specified total number of edges are assigned to nodes completely at random, thus yielding a Poisson degree distribution across the network.

The structures of human contact networks undoubtedly play a crucial role in the transmission of diseases (Pastor-Satorras & Vespignani 2001; Newman 2002; Barthélemy *et al.* 2004, 2005; Meyers *et al.* 2005; Ferrari *et al.* 2006). For example, epidemiologists have long realized that the epidemic threshold, the critical value for the infection rate above which a disease may spread and persist, decreases as the standard deviation of the degree distribution increases (Hethcote & Yorke 1984; Anderson & May 1992). The obvious question is thus: what are the structures of real-world contact networks?

The field has focused particularly on scale-free random networks (May & Lloyd 2001; Dezso & Barabasi 2002; Pastor-Satorras & Vespignani 2002), based on the apparent ubiquity of such networks in natural and human-made systems and a (limited) set of studies of epidemiologically relevant contact patterns (Liljeros *et al.* 2001). These networks are characterized by the presence of hosts with anomalously high numbers of potential disease-causing contacts, called super-spreaders, which have important epidemiological implications (Shen *et al.* 2004; Lloyd-Smith *et al.* 2005). With large or infinite variance in degree, scale-free networks can have exceedingly low or non-existent epidemic thresholds; this means that even the sparsest networks are highly vulnerable to epidemics. Despite recent popularity, however, it is not clear that realistic epidemiological networks are generally scale-free, and thus whether they deserve so much attention in epidemiology. Here, we address this by characterizing the structures of several real-world networks.

We limit our discussion to random networks with arbitrary degree distributions, including random networks with regular, Poisson, exponential and scale-free degree distributions (figure 1). These classes of networks have been well studied with respect to the spread of epidemics and are representatives of a spectrum of network structures. We focus exclusively on the epidemiological impact of the degree distribution, although other network characteristics such as clustering (Watts & Strogatz 1998; Keeling 1999; Moore & Newman 2000; Petermann & Rios 2004) and degree correlations (Boguna *et al.* 2003) are also important.

We will also focus exclusively on static networks, i.e. networks in which contacts are assumed to be fixed during the infectious period of an individual. The permanence of contacts captured by static networks offers a more realistic model of human contact behaviour than that by traditional epidemiological models. However, a recent study (Volz & Meyers submitted) suggests that static networks may only be an approximate model for diseases that spread slowly relative to the rate at which individuals change the numbers and identities of their contacts.

3. A NETWORK INTERPRETATION OF HOMOGENEOUS-MIXING MODELS

Traditional epidemiological models implicitly assume that contact patterns are highly homogeneous. The most basic model in epidemiology is the SIR compartmental model for the 'simple' epidemic without demographic processes of birth and death (Kermack & McKendrick 1991; Anderson & May 1992). This model describes the dynamics of a single epidemic, during the course of which each individual is in one of three disjoint states (or compartments) at any given time: not yet infected and susceptible to disease (S); infected and infectious (I); or recovered and neither able to spread disease nor be reinfected (R). The number of individuals in each of these classes is described by the following set of differential equations:

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -\lambda S, \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \lambda S - \gamma I, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I, \end{aligned} \tag{3.1}$$

where λ is the rate at which susceptible individuals become infected (i.e. the *force of infection*) and γ is the recovery rate for infected individuals. (The constant recovery rate, γ, yields an exponential distribution of infectious periods.)

The parameter λ can be thought of as a composite of three factors: $\lambda = \alpha \iota_t \tau$, where α is the number of individuals with which a susceptible individual has effective contact; $\iota_t = I(t)/N$ is the proportion of contacts that are infectious; and τ is the per contact rate at which disease is transmitted between an infectious and susceptible individual (Keeling & Eames 2005). This model assumes homogeneous-mixing among individuals. Taken literally, the assumption means that at any point in time, every susceptible individual has an equal probability of contacting every other individual in the population. Taken more loosely, it is sufficient for all susceptible individuals to have comparable contact patterns (α) and encounter infected individuals at a rate corresponding to the overall prevalence ($\iota_t = I(t)/N$).

We suggest that the standard equations given above can be approximately mapped to a network model in which all individuals have identical numbers of contacts—a regular random network. Consider a regular random network with homogeneous degree k. Suppose that disease spreads from infected nodes to susceptible contacts with a rate of τ'. Because contacts are random, we can assume that the fraction of contacts that are infected is equal to the fraction of infected individuals in the network as a whole, or $\iota_t = I(t)/N$. Then, the force of infection on any given susceptible individual will be $\lambda' = k\iota_t\tau'$, which is equal to that given by the homogeneous-mixing compartmental model with $\alpha = k$ and $\tau = \tau'$. This mapping is only approximate because the two models view effective contact differently. In compartmental models, the identity of a contact is determined randomly and instantaneously for each transmission event; in static network models, on the other hand, the identities of contacts are chosen randomly but remain fixed for the length of an individual's infectious period. The models thus yield different values for the basic reproductive ratio and expected equilibrium values (Keeling & Grenfell 2000).

Some authors have argued that the homogeneous-mixing assumption translates into a network where every individual is connected to every other individual in the population (a complete network; Aparicio & Pascual 2007; Roy & Pascual 2006). In fact, a complete network is just a regular random network in which each individual contacts all $N-1$ other individuals, and thus is one of many valid (regular) network interpretations of the equations above. The force of infection in a complete network is $\lambda' \approx (N-1)\iota_t\tau'$. For most populations, however, it is unrealistic to imagine that individuals are in constant contact with all other individuals. Thus, it is more realistic to interpret the edges in the complete network as 'possible' contacts and the transmission term τ' as a combined effective contact and transmission rate.

Figure 2*a*,*b* demonstrates that, indeed, the dynamics predicted by the homogeneous-mixing compartmental model are a close approximation to stochastic simulations of disease transmission through a regular random network. The predicted final size of an outbreak (figure 2*a*) is shown as a function of T, the per contact probability of transmission integrated over the entire infectious period of an infected individual. We generated these and all other random networks described in §5 using the configuration model (Molloy & Reed 1995). The stochastic epidemiological simulations are based on a discrete-time, chain binomial, SIR model (Bailey 1957). (See appendix A in the electronic supplementary material for further details.)

4. HETEROGENEITY IN EMPIRICAL CONTACT NETWORKS

The homogeneous-mixing assumption is reasonable when contact patterns in populations are random and homogeneous; that is, they resemble a regular random network. The contact patterns in real populations, however, may be more heterogeneous than assumed by the simple models. Figure 2*c*–*h* illustrates that the simple models become inadequate as contact patterns become more variable. The predictions of the homogeneous-mixing model are more reasonable for a Poisson random network (figure 2*c*,*d*)—which is more variable than a regular random network, but less so than an exponential or scale-free network—than they are for networks with more variation (figure 2*e*–*h*). Here, we quantify the heterogeneity found in realistic contact networks and its epidemiological implications.

4.1. Statistical analysis of realistic networks

Using datasets from the literature, we characterize the variation found in real-world contact networks. In particular, we determine the shape of the degree distributions in six empirical (or semi-empirical) social networks. Network characterizations are often based on graphical methods for fitting data to theoretical

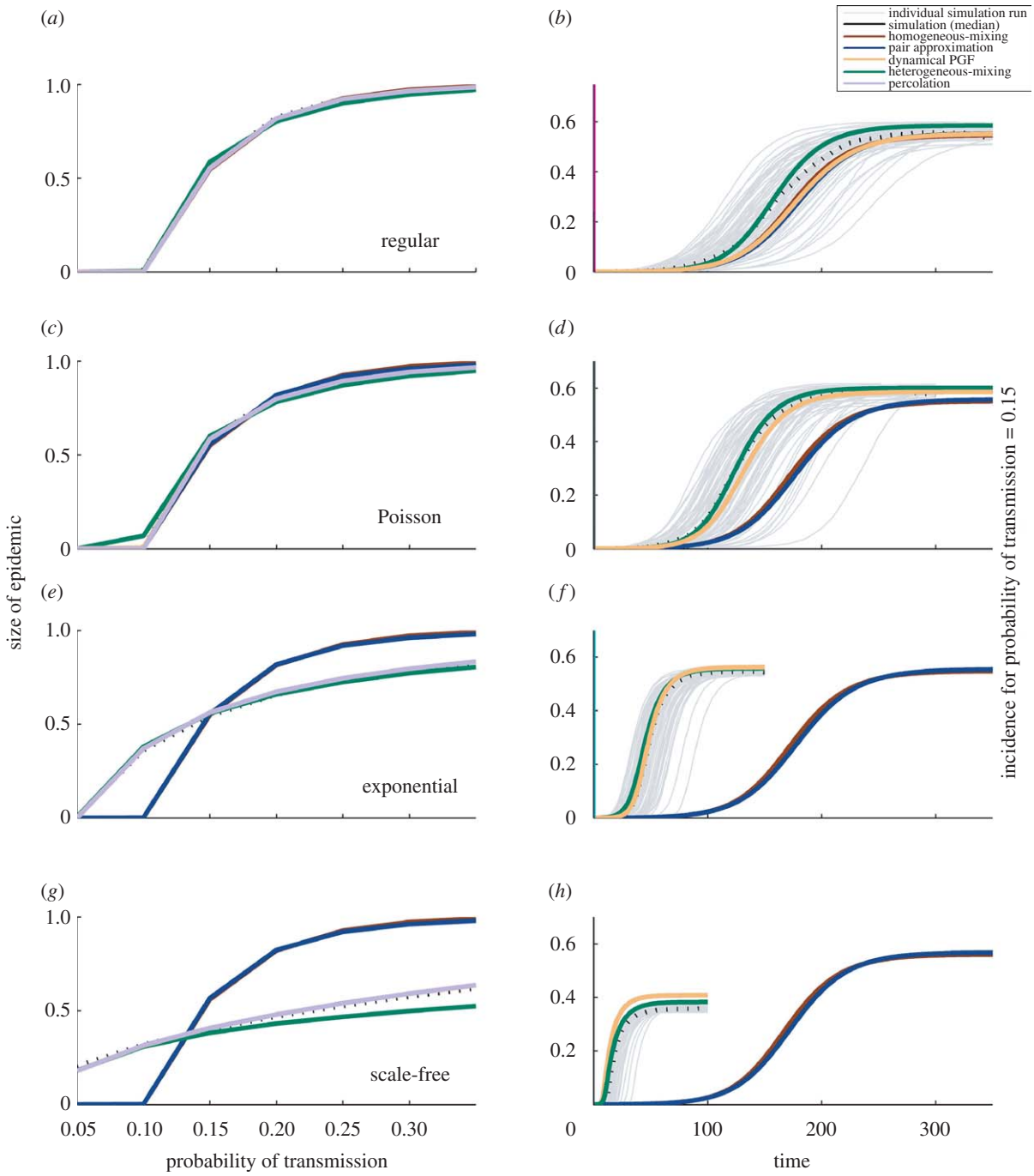

Figure 2. A comparison of the homogeneous-mixing compartmental and network models on various random networks. The networks each have 10 000 nodes and a mean degree of 10, with regular, Poisson, exponential and scale-free degree distributions, respectively. (*b*,*d*,*f*,*h*) Grey lines, individual simulation runs; dotted black line, the median of values from the simulations. The homogeneous-mixing model is as described in §3. The four network-based models are the pair approximation model (Keeling 1999), percolation model (Newman 2002), heterogeneous-mixing model (Moreno *et al.* 2002) and dynamical PGF model (Volz in press). In (*a*,*c*), all curves overlap. In (*b*), curves for homogeneous-mixing, pair approximation and dynamical PGF overlap. In (*d*–*h*), curves for homogeneous-mixing and pair approximation overlap. In (*e*,*g*), curves for dynamical PGF and percolation completely overlap. (Percolation does not provide dynamical predictions and is thus not graphed in (*b*), (*d*), (*f*) or (*h*).)

distributions. This approach, however, can be flawed, especially if the data are transformed to a log–log scale (Goldstein *et al.* 2004). For each of our datasets, we evaluate four one-parameter candidate distributions (Poisson, exponential, pure power law and truncated power law) using maximum-likelihood estimation (MLE) to fit the distribution parameters. MLE avoids the problems of visual/graphical fitting methods and more accurately estimates the distribution parameters. We then use the Akaike information criterion (AIC) to select the most appropriate distribution for the data (see appendix B in the electronic supplementary material for further methodological details).

Our datasets are chosen from the limited set of empirical contact networks available in the literature and include several different scales and types of contacts. The first is an urban contact network model based on demographic information for the city of Vancouver, British Columbia (Meyers *et al.* 2005). The second is another urban contact network model

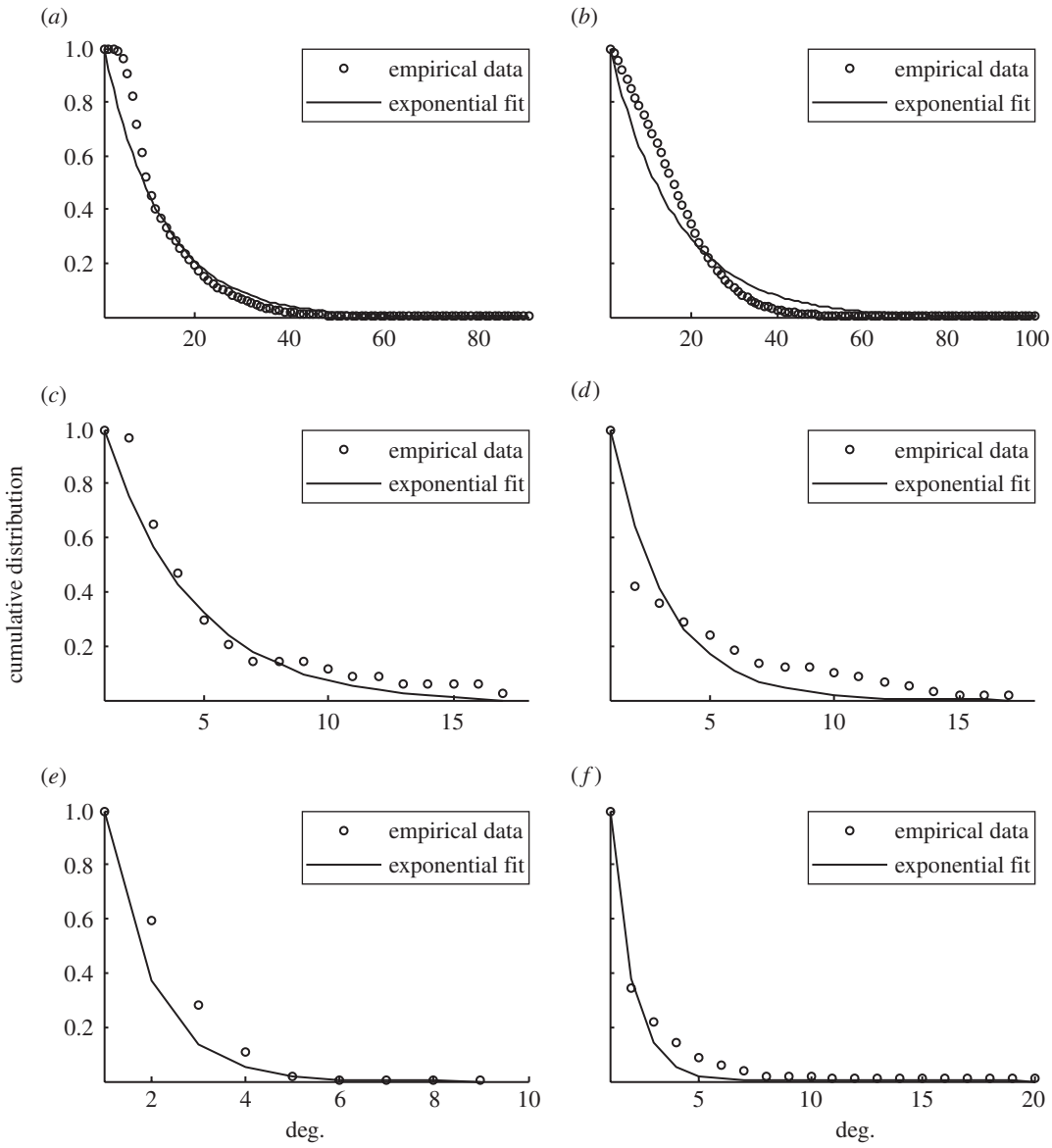

Figure 3. Statistical fitting of empirical datasets: (*a*) Vancouver urban network (Meyers *et al.* 2005), (*b*) Portland urban network (Eubank *et al.* 2004; Del Valle *et al.* 2006), (*c*) Zachary Karate Club network (Zachary 1977), (*d*) Atlanta high school syphilis network (Rothenberg *et al.* 1998), (*e*) Midwest town adolescents network (Bearman *et al.* 2004), and (*f*) Colorado Springs risk network (Potterat *et al.* 2002). All datasets fit best to the exponential distribution with the parameter values given as follows: (*a*) $\lambda = 13.11 \pm 9.9 \times 10^{-3}$, (*b*) $\lambda = 15.94 \pm 2.2 \times 10^{-6}$, (*c*) $\lambda = 4.07 \pm 7.8 \times 10^{-3}$, (*d*) $\lambda = 2.72 \pm 1.4 \times 10^{-2}$, (*e*) $\lambda = 1.46 \pm 2.8 \times 10^{-2}$, and (*f*) $\lambda = 1.47 \pm 2.8 \times 10^{-2}$ (along with standard error in the parameter estimate).

developed for the city of Portland, Oregon based on surveys and other demographic and geographical data (Eubank *et al.* 2004; Del Valle *et al.* 2006). The third dataset describes social ties (which may be used as a proxy for disease-causing contacts) among members of a university karate club (Zachary 1977). The final three studies provide sexual contact patterns within different adolescent populations (Rothenberg *et al.* 1998; Potterat *et al.* 2002; Bearman *et al.* 2004).

Sampling bias in data collection is an important concern in the study of empirical contact networks, and its impact has yet to be understood for epidemiological applications. Preliminary work by Stumpf *et al.* (2005) has shown that scale-free networks are especially prone to sampling errors. In fitting the datasets above to ideal distributions, we have not made any additional adjustments to correct for possible sampling biases (beyond those made in the original studies), which may obscure the true structure of the population.

Our analysis suggests that all six populations fit an exponential degree distribution best, based on the AIC (figure 3). We note that this is the best choice only among the four single-parameter distributions considered, and that other multiple-parameter distributions like gamma or beta distributions might yield even better approximations of the data. Although the exact shape of these distributions may be uncertain given the small sample sizes, our analysis suggests generally that the variability found in these realistic contact networks appears to lie somewhere between the homogeneity assumed by homogeneous-mixing models and the high heterogeneity of scale-free networks. Recently, Amaral *et al.* (2000) found similarly that adolescent friendship networks have exponentially shaped or Gaussian degree distributions.

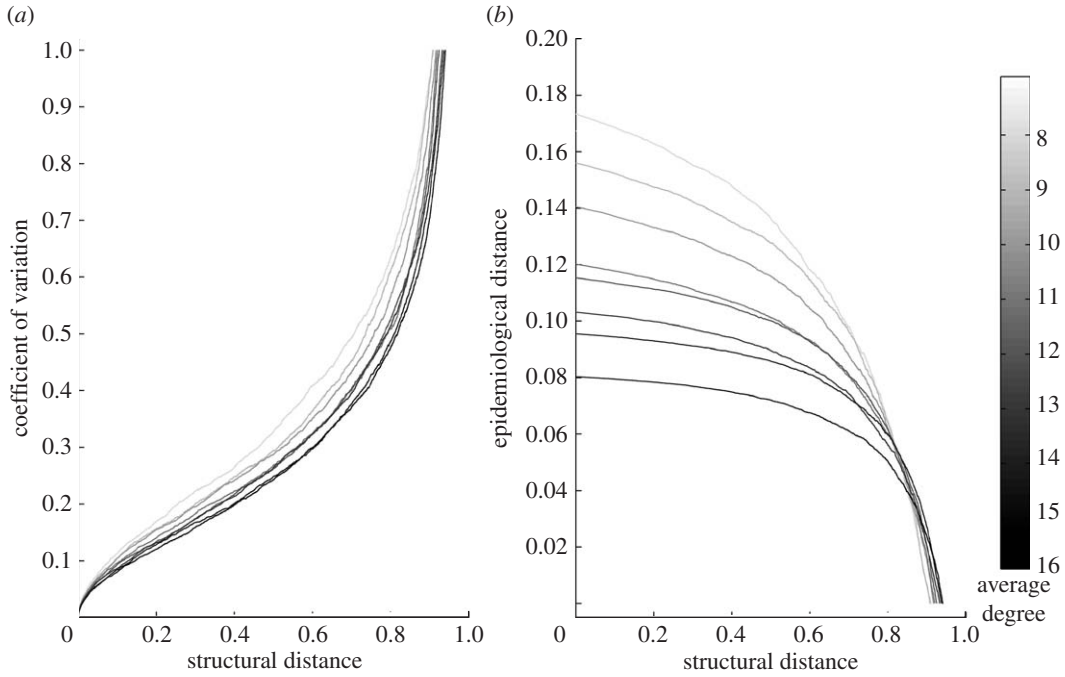

Figure 4. (*a*) Structural distance (measured as the probability of an edge to be rewired) versus the coefficient of variation (CV). (*b*) Structural distance versus epidemiological distance, measured as relative difference in predictions for the final size (S) of epidemic for the current network and the target (exponential) network $S_{\text{current}} - S_{\text{target}}/S_{\text{target}}$. Both plots include several lines corresponding to different mean degrees, from 8 to 16 (from grey to black, respectively). The number of rewirings required to reach a CV of 1 increases with the mean of the network. All epidemiological calculations in (*b*) assume the probability of transmission $T=0.25$.

4.2. Quantifying deviation from assumptions of homogeneity

Given the nearly exponential shape of the empirical contact networks we have considered, we next quantify the epidemiological distance between exponential networks and the homogeneous-mixing model (equivalently, a regular random network) using a structural approach. Starting with a regular random network, we use a greedy rewiring procedure to gradually generate an exponentially distributed network. In particular, we iteratively select a random edge in the network, and change its destination to a new node (chosen in proportion to its degree). This process reduces the degree of the original destination node and increases the degree of the new destination node (which already had high degree), and thus progressively increases the variance in the degree distribution. (Details on the rewiring algorithm can be found in appendix C in the electronic supplementary material.)

We monitor the structural evolution of the network during the rewiring process in terms of the coefficient of variation (CV) in degree—the standard deviation in degree divided by the mean degree. We have observed that when the CV reaches 1, which is the CV of an exponential degree distribution, the modified networks have approximately exponential degree distributions (appendix C in the electronic supplementary material).

In figure 4, we show the impact of rewiring on both the structure of the network (figure 4*a*) and the resulting epidemiological dynamics (figure 4*b*). For any given network created by rewiring, we define its *structural distance* from the original regular network as the probability that a randomly selected edge has been rewired. This is estimated by $1-(1-1/m)^r$, where m is the number of edges in the network and r is the number of rewiring events so far. We show results for networks with mean degrees varying from $\langle k\rangle = 8$ to 16, which spans part of the range of contact patterns observed in the realistic networks discussed in §4.1. Figure 4*a* illustrates that the structural progression towards exponential networks slows as the mean degree of the network increases.

We define the *epidemiological distance* to the target (exponential) network as the relative discrepancy between the expected size of an epidemic on the current network and the expected size of an epidemic on the target network, based on stochastic simulations. We use these quantities to characterize the rate at which epidemiological behaviour changes as a function of structure (contact heterogeneity). Figure 4*b* illustrates the epidemiological consequences of rewiring for a particular probability of transmission ($T=0.25$). The y-intercept values in figure 4*b* (structural distance $=0$) indicate the amount of error realized if one uses a homogeneous-mixing model to make predictions on an exponential network. For example, in a population with a mean degree of 8, the homogeneous-mixing model has a 17% error in the prediction for final size of epidemic; while for a network with a mean degree of 16 (such as the urban contact network for Portland, Oregon discussed in §4.1), the error is approximately 8%. In general, for a given probability of transmission, the discrepancy between the homogeneous-mixing model and the true epidemiology decreases as the mean degree increases. Thus, for highly connected exponential networks, the homogeneous-mixing compartmental models may offer reasonable approximations.

As rewiring progresses, the degree distribution of the network approaches an exponential distribution, and the epidemiological behaviour of the network

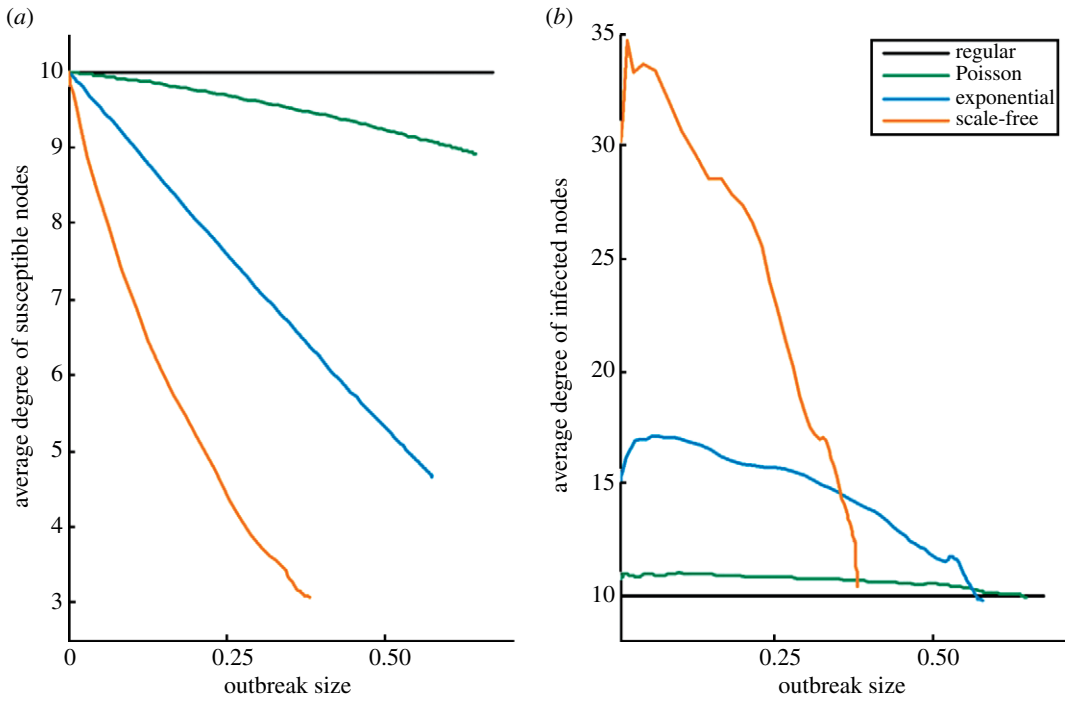

Figure 5. The average degree among currently (*a*) susceptible and (*b*) infected/infectious nodes for various networks. Each network has 10 000 nodes having a regular, Poisson, exponential or scale-free degree distributions, each with a mean of 10. Values are averaged across 50 simulation runs. The x-axis gives the cumulative incidence $(1-S/N)$.

approaches that of an exponential network, at first slowly and then rapidly. The shape of the curves in figure 4*b* indicates that essentially the entire network must be restructured before the epidemiological consequences begin to resemble those of an exponential network. In other words, epidemiological models (beyond those that make the homogeneous-mixing assumption) need to incorporate almost all of the variation of the true network in order to reduce the error in the epidemiological predictions even slightly.

5. ANALYTICAL APPROACHES FOR INCORPORATING HETEROGENEITY

As the tail of a network's degree distribution grows, i.e. as the variability in the number of contacts increases, disease dynamics increasingly differ from those predicted by homogeneous-mixing models in two respects. First, early in an outbreak, the probability of a contact between a susceptible and infected individual is *higher* than expected because outbreaks are initially biased towards individuals who have high numbers of contacts and are thus epidemiologically vulnerable. Second, late in an outbreak, the reverse occurs, and disease-causing contacts are fewer than expected because the newly infected and remaining susceptible populations tend to have low numbers of contacts. Figure 5 illustrates this changing distribution of contacts in the susceptible and infectious populations over the course of an outbreak. As the disease spreads, not only does the average number of contacts per infected individual change, but also the variability in the number of contacts decreases. Consequently, the homogeneous-mixing models underestimate disease burden early in the outbreak and overestimate it towards the end, as illustrated in figure 2*d*,*f*,*h*.

5.1. Network and other individual-based models

Incorporating these heterogeneities into analytical models of disease spread has proved to be a critical challenge. Differential equations-based models enjoy the advantages of being tractable, yielding to sensitivity analysis and providing temporal behaviour, but they often lack attention to detail in the underlying process. Individual-based models allow individual-level heterogeneities to be taken into account and can prove to be very useful in making public health decisions, but can often be intractable. Despite the challenges, many powerful analytical approaches have been developed in recent years to make predictions of disease spread in heterogeneous populations. We review some of these approaches here.

Percolation theory methods are based on generating functions, and only require the degree distribution of the network and the average transmissibility T of the pathogen, i.e. the probability that an infected individual will transmit disease to a susceptible contact during his or her infectious period. These methods are very general and mathematically tractable. They provide excellent final-state predictions, but do not predict the dynamics of an outbreak (Moore & Newman 2000; Newman 2002; Meyers *et al.* 2003, 2005, 2006). Pair approximation methods (Keeling 1999) are based on a differential equation model of counting the number of pairs of individuals in each disease class. These methods are particularly useful for networks with clustering or spatial networks. There are many useful extensions of pair approximation methods for networks with various structures (e.g. Eames & Keeling 2002), but we consider only the simplest of these models in the analysis below. Moreno *et al.* (2002) and Pastor-Satorras & Vespignani (2002) have developed a dynamical framework based on

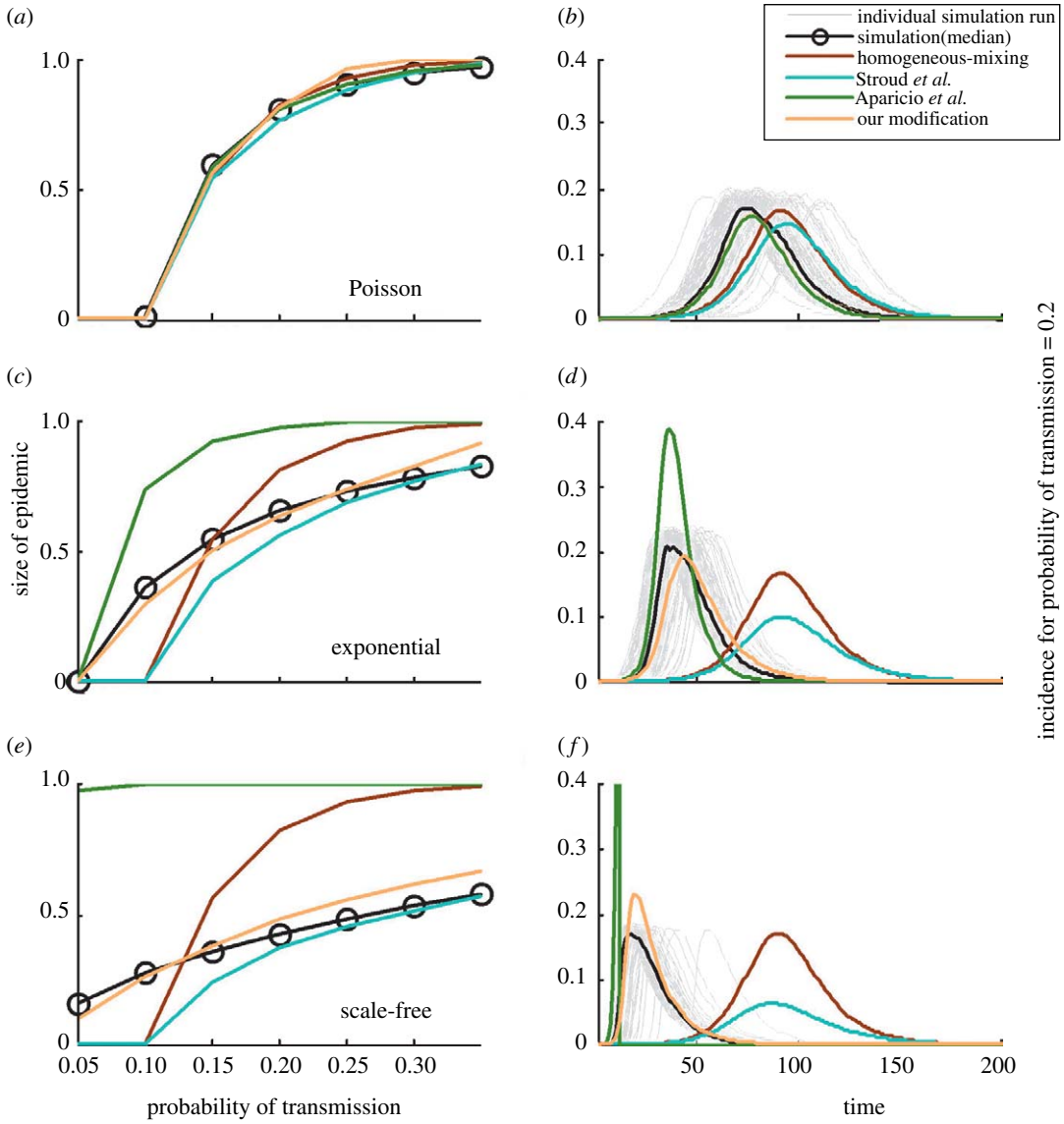

Figure 6. Epidemiological predictions with 'modifications' to the homogeneous-mixing compartmental model on three classes of networks. The networks each have 10 000 nodes and a mean degree of 10, with Poisson, exponential and scale-free degree distributions, respectively. (b,d,f) Grey lines, individual simulation runs; black line, the median of values from the simulations. 'Homogeneous-mixing' refers to the homogeneous-mixing compartmental model described in §3. Predictions from the modifications by Aparicio & Pascual (2007) and Stroud *et al.* (2006) are shown. 'Our modification' refers to the modified force of infection parameter (λ_t') described in §5.2.

a system of degree-based differential equations. These models track numbers of individuals in each disease state for each degree class (there are $3k$ differential equations in the system for an SIR model, where k is the number of unique degree classes). Finally, Volz (in press) has developed a powerful probability generating function (PGF)-based differential equation model that accurately tracks global epidemiological dynamics in complex random networks (we give details of each of these approaches in appendix D in the electronic supplementary material).

In figure 2, we compare the predictions of these four network-based methods for four different types of random networks: regular; Poisson; exponential; and scale-free. The networks are generated using the configuration model (Molloy & Reed 1995) and the parameters for the degree distributions are chosen so that each network has an identical average degree of 10. In particular, let p_k be the probability that a randomly chosen node has degree k. All nodes in the regular network have degree $k=10$; the Poisson degree distribution is given by $p_k = \mathrm{e}^{-\theta}\theta^k/k!$, where $\theta=10$; the exponential degree distribution is given by $p_k = (1-\mathrm{e}^{-1/\theta})\mathrm{e}^{-1/\theta(k-1)}$, where $\theta=9.49$; and the scale-free degree distribution is given by $p_k = k^{-\theta}/\zeta(\theta)$, where $\theta=1.875$ and $\zeta(\theta)$ is the Riemann zeta function. We thereby hold network density constant and investigate only the effects of contact heterogeneity on disease dynamics.

The predictions from these models are compared with stochastic simulations of disease transmission on the same networks in terms of both the total burden of morbidity in the population (figure 2*a*,*c*,*e*,*g*, dotted black line) and the disease incidence dynamics (figure 2*b*,*d*,*f*,*h*, grey lines). Accurate predictions of these quantities are critical to efficient and timely implementation of public health measures. Generally, the homogeneous-mixing model does not perform as well as the other methods on complex random graphs. The percolation,

heterogeneous-mixing and the dynamical PGF-based methods generally perform well, but may require more information (namely the degree distribution) than the other methods. Recall, however, that the realistic networks described above were reasonably approximated by single-parameter exponential degree distributions. Whenever a network closely resembles a random network with an idealized, low-parameter degree distribution, these methods do not require many parameters, if any more than the corresponding compartmental models.

5.2. Modified compartmental models

We have argued that the homogeneous-mixing compartmental model essentially assumes that contact patterns within a population form a regular random network, and shown that real-world contact patterns often exhibit more heterogeneity. The homogeneous-mixing compartmental model fails to make accurate predictions for such networks because it does not account for the evolving structure of the population (figure 5). In particular, when there is significant variability in contact rates, contacts between infected and susceptible individuals will initially be more frequent than predicted by homogeneous-mixing models.

Although there are several powerful network-based analytical methods for predicting disease transmission on complex networks, it would still be desirable to develop a method that works within the simple compartmental framework (Levin & Durrett 1996; Diekmann *et al.* 1998). If successful, this would help bridge the conceptual gaps between the new and old methods and offer an appealing alternative to researchers already comfortable with this methodology. Several epidemiologists have taken steps in this direction (Severo 1969; Liu *et al.* 1987; Heathcote & Nicholls 1990; Hochberg 1991; Keeling 2005; Aparicio & Pascual 2007; Stroud *et al.* 2006; Roy & Pascual 2006).

The most recent effort has been the most direct one: Aparicio & Pascual (2007) have modified the homogeneous-mixing compartmental model with a network-based approximation of the reproductive ratio, R_0, and have demonstrated the success of the approach on Poisson and small-world random networks. Although effective for these particular networks, the modification does not capture the evolution of contact patterns that occurs over the course of an outbreak. Figure 6 compares this approach with the homogeneous-mixing compartmental model. Aparicio and Pascual do not actually discuss adaptations of their model to exponential or scale-free networks. To evaluate its performance on such networks, we have inserted analytical estimates of R_0, based on the specific structures of these networks (Meyers *et al.* 2005). That is, $R_0 = (\tau/\tau+\gamma)(\langle k^2\rangle - \langle k\rangle/\langle k\rangle)$, where τ is the average probability of infection per contact and time step; γ is the recovery probability; and $\langle k\rangle$ and $\langle k^2\rangle$ are the average degree and average squared degree in the networks, respectively. As seen in figure 6*c–f*, the modification based on the expected number of secondary cases early in the outbreak gives a gross overestimate of epidemiological predictions.

A different class of approaches to bridge the gap between homogeneous and individual-based models uses epidemic data. Keeling (2005) suggests modifying the transmission term to be a function of time, $\beta(t)$, and fitting this time-dependent term to case-reporting data. Keeling has shown that this approach works well, given sufficient epidemiological data. Using a model similar to those introduced by Severo (1969), Liu *et al.* (1987) and Hochberg (1991), Roy & Pascual (2006) suggest modifying the infection term of the homogeneous-mixing compartmental model to be $\tau k S^p I^q$, where the 'heterogeneity parameters' k, p and q are estimated via least-squares fitting to simulation data. They demonstrate the success of this approach on small-world networks. Similarly, Stroud *et al.* (2006) modify the term representing the proportion of susceptibles in the homogeneous-mixing model with an empirical exponent $(S/N)^v$, where v is estimated from simulation data. They successfully demonstrate this approach on several urban contact networks (Stroud *et al.* 2006). We choose to evaluate the Stroud model as a representative of this class of modifications (figure 6).

Figure 6 shows that neither of the tested modifications succeeds in capturing the disease dynamics across all network classes and probabilities of transmission. These efforts to account for contact heterogeneity seem somewhat ad hoc, and not justified from 'first principles'. A flexible modification to the simple framework must be able to capture the evolving distributions of contacts illustrated in figure 5. During the course of an epidemic, the epidemiologically active portion of the network (individuals who are either susceptible or infectious) shrinks and changes in contact structure. The nature of this change fundamentally depends on the initial network structure. The homogeneous-mixing compartmental model takes into account the changing *densities* of the epidemiologically active individuals in the population (changing values of S and ι_t), but not the changing *distribution of contacts* among epidemiologically active individuals (constant α).

Looking closer at figure 5, we observe that, in heterogeneous networks, the average degree of infected nodes is initially significantly higher than the average degree in the network as a whole (all four networks have an average degree of 10). This is because high-degree nodes are more likely to be infected than low-degree nodes. This value decreases as the epidemic progresses and high-degree nodes recover from disease and thereby are removed from the epidemiologically active portion of the network. The average degree among susceptible nodes likewise decreases owing to the removal of high-degree nodes. This figure also sheds light on the mapping of the homogeneous-mixing compartmental model onto the regular random network model as discussed in §3. The contact structure of the regular random graph cannot change since all nodes have identical numbers of contacts.

In theory, the simple compartmental model can be modified to capture this structural evolution. We recall that the equation describing the dynamics of infected individuals in the homogeneous-mixing compartmental model is

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \lambda S - \gamma I,$$

where $\lambda = \alpha \iota_t \tau$. Our discussion above suggests that α, which was originally defined as the number of contacts for a typical node (i.e. the average degree in the original network $\langle k \rangle$), should be modified to give the average number of contacts among *currently* susceptible nodes, as given by

$$\alpha'_t = \langle k \rangle_{S(t)}.$$

We must also modify ι_t from the fraction of currently infected individuals to the fraction of all edges (contacts) that lead to infected individuals. The total number of edges leading to infected individuals at time t is $I(t)\langle k \rangle_{I(t)}$, where $\langle k \rangle_{I(t)}$ is the average degree among currently infected nodes, and the number of all contacts in the network is $N\langle k \rangle$, where $\langle k \rangle$ is the average degree in the original network. Thus, we define

$$\iota'_t = \frac{I(t)\langle k \rangle_{I(t)}}{N\langle k \rangle} = \iota_t \frac{\langle k \rangle_{I(t)}}{\langle k \rangle}.$$

With these modifications to the model parameters, we can redefine the force of infection as

$$\lambda'_t = \alpha'_t \iota'_t \tau = \frac{\langle k \rangle_{S(t)} \langle k \rangle_{I(t)}}{\langle k \rangle} \iota_t \tau.$$

This modification to the standard compartmental model (replacing the force of infection term λ with λ'_t) is both flexible and theoretically motivated. The exact form of λ'_t, however, depends on the contact structure of the population and can be found with one of two methods. One could use analytical methods borrowed from network modelling to estimate (or calculate exactly) the function λ'_t. This would mean, however, that an entire set of network equations would have to be embedded within the compartmental framework, defeating the purpose of a simple modification. Another option would be to simulate exactly a full epidemic on the contact network, and estimate λ'_t by fitting linear or quadratic functions to the values of $\langle k \rangle_{S(t)}$ and $\langle k \rangle_{I(t)}$ as a function of outbreak size from the simulation data. Figure 6 shows that this option indeed performs better than existing modifications, but is cumbersome in comparison with the latest generation of network models. Thus, we do not advocate its general use, but present it simply to provide insight into the mechanisms of infection during an epidemic.

6. CONCLUSION

Homogeneous-mixing models have had a long history of success (Anderson & May 1992) and continue to produce valuable results. Until recently, individual-based models have been considered computationally prohibitive and unnecessary, given the flexibility of the compartmental framework to analytically divide hosts into multiple demographic classes. With the recent development of several powerful analytical approaches to model disease spread through heterogeneous populations, however, we must ask: what have the simplest models been missing, if anything and which models should we use in the future?

The answer to the first question is not as bleak as one might fear. Epidemiologists have recently focused on scale-free contact networks made up of a chaste majority and a small, highly promiscuous minority. Human contact patterns indeed exhibit more heterogeneity than assumed by homogeneous-mixing models, but they do not appear generally to be scale-free. The diverse populations we considered above have exponentially distributed contact patterns (figure 3). The epidemiological behaviour of exponential networks is typically much closer to the predictions of the homogeneous-mixing models than that of scale-free networks, although this depends on the transmission rate of the pathogen (figure 2).

The answer to the second question, in some cases, is a matter of taste. For many studies, there is a single population (and thus a contact network) under consideration. If the network is close to homogeneous, the homogeneous-mixing compartmental model is a reasonable choice, although network models are equally tractable, and can be reduced to a few parameters by assuming an idealized approximation for the degree distribution (figure 2). If the population is heterogeneous, but falls into a few specific classes of networks, there may be simple modifications to the homogeneous-mixing compartmental framework that perform quite well (e.g. Aparicio & Pascual 2007), and again one has a choice of frameworks (figure 6). If, however, the network is exponential or scale-free, or perhaps is not well approximated by any of these common distributions, then neither the homogeneous-mixing model nor the modifications can adequately capture the evolving structure of the host network (figure 5).

We have not considered several classes of more complex compartmental models that incorporate various sources of heterogeneity in the host population. Through additional parameters, these may perform better than the simple models on complex host networks, or be more easily adapted to do so. For example, Hethcote and Yorke successfully modelled the spread of gonorrhoea by dividing populations into several homogeneously mixed groups based on gender and sexual activity (Hethcote & Yorke 1984). Anderson and May evaluated vaccination programmes for childhood infections using age-structured models, in which the transmission term is replaced by a 'who acquires infection from whom' matrix of transmission rates (Anderson & May 1985). Bjornstad and Grenfell have developed the time-series SIR (TSIR) framework to model seasonal changes in contact patterns that influence the spread of measles (Bjørnstad *et al.* 2002; Grenfell *et al.* 2002). Ball and colleagues introduced a general patch model in which hosts mix at high rates locally and at low rates globally (Ball *et al.* 1997). Although these approaches have been highly successful, we note that several of them require an *a priori* categorization of individuals into epidemiological type classes (e.g. by age, sex or location). In general, it may be difficult to identify epidemiologically meaningful groupings, particularly for a newly emerging disease.

When contacts are heterogeneous, it thus makes sense to consider the growing toolkit of tractable network-based methods that explicitly consider individual-level contact patterns and can predict the changing structure of a contact network as disease spreads through it. These methods may appear intimidating, but are actually

intuitive upon closer inspection. They enable straightforward mathematical calculations of disease transmission dynamics through non-standard populations, and allow detailed demographic predictions that can serve as vital input to public health decisions.

The authors thank Richard Rothenberg and Sara Del Valle for sharing data; Shashank Khandelwal for algorithm implementation; Erik Volz for technical advice and stimulating discussions; and three anonymous referees for their valuable input. S.B. acknowledges the support of the NASA-Jenkins Fellowship Program and L.A.M. acknowledges grant support from the James S. McDonnell Foundation.

REFERENCES

Addy, C. L., Longini, I. M. & Harber, M. 1991 A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* **47**, 961–974. (doi:10.2307/2532652)

Albert, R., Jeong, H. & Barabasi, A.-L. 1999 The diameter of the World Wide Web. *Nature* **401**, 130–131. (doi:10.1038/43601)

Amaral, L. A. N., Scala, A., Barthelemy, M. & Stanley, H. E. 2000 Classes of small-world networks. *Proc. Natl Acad. Sci. USA* **97**, 11 149–11 152. (doi:10.1073/pnas.200327197)

Anderson, R. M. & May, R. M. 1985 Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *J. Hyg. (Lond.)* **94**, 365–436.

Anderson, R. M. & May, R. M. 1992 *Infectious diseases of humans: dynamics and control*. New York, NY: Oxford University Press.

Aparicio, J. P. & Pascual, M. 2007 Building epidemiological models from R_0: an implicit treatment of transmission in networks. *Proc. R. Soc. B* **274**, 505–512. (doi:10.1098/rspb.2006.0057)

Bailey, N. T. J. 1957 *The mathematical theory of epidemics*. London, UK: Griffin.

Ball, F., Mollison, D. & Scalia-Tomba, G. 1997 Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–49. (doi:10.1214/aoap/1034625252)

Barabási, A.-L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)

Barbour, A. & Mollison, D. 1990 Epidemics and random graphs. In *Stochastic processes in epidemic theory* (eds J. P. Gabriel, C. Lefevre & P. Picard), pp. 86–89. New York, NY: Springer.

Barthélemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. 2004 Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178 701. (doi:10.1103/PhysRevLett.92.178701)

Barthélemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. 2005 Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J. Theor. Biol.* **235**, 275–288. (doi:10.1016/j.jtbi.2005.01.011)

Bearman, P. S., Moody, J. & Stovel, K. 2004 Chains of affection. *Am. J. Sociol.* **110**, 44–91. (doi:10.1086/386272)

Bjørnstad, O. N., Finkenstädt, B. & Grenfell, B. T. 2002 Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* **72**, 169–184.

Boguna, M., Pastor-Satorras, R. & Vespignan, A. 2003 Epidemic spreading in complex networks with degree correlations. In *Statistical mechanics of complex networks* (eds J. M. Rubi *et al.*), pp. 127–147. Berlin, Germany: Springer.

Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K., Lee, I. & Marcotte, E. M. 2004 Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.* **14**, 292–299. (doi:10.1016/j.sbi.2004.05.003)

Centers for Disease Control and Prevention 2000 Measles outbreak—Netherlands April 1999–January 2000. *Morb. Mortal. Wkly Rep.* **49**, 299–303.

Del Valle, S. Y., Stroud, S. Y., Smith, J. P., Mniszewski, S. M., Riese, J. M. & Sydoriak, S. J. 2006 *EpiSimS: epidemic simulation system*. Los Alamos, NM: Los Alamos National Laboratory.

Dezso, Z. & Barabasi, A. L. 2002 Halting viruses in scale-free networks. *Phys. Rev. E* **65**, 055103. (doi:10.1103/PhysRevE.65.055103)

Diekmann, O., De Jong, M. C. M. & Metz, J. A. J. 1998 A deterministic epidemic model taking account of repeated contacts between the same individuals. *J. Appl. Prob.* **35**, 448–462. (doi:10.1239/jap/1032192860)

Eames, K. T. D. & Keeling, M. J. 2002 Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl Acad. Sci. USA* **99**, 13 330–13 335. (doi:10.1073/pnas.202244299)

Eubank, S., Guclu, H., Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z. & Wang, N. 2004 Modeling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184. (doi:10.1038/nature02541)

Faloutsos, M., Faloutsos, P. & Faloutsos, C. 1999 On power-law relationships of the internet topology. In *Proc. Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications*.

Ferrari, M. J., Bansal, S., Meyers, L. A. & Bjørnstad, O. N. 2006 Network frailty and the geometry of herd immunity. *Proc. R. Soc. B* **273**, 2743–2748. (doi:10.1098/rspb.2006.3636)

Giot, L. *et al.* 2003 A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736. (doi:10.1126/science.1090289)

Goldstein, M. L., Morris, S. A. & Yen, G. G. 2004 Problems with fitting to the power-law distribution. *Eur. Phys. B* **41**, 255–258. (doi:10.1140/epjb/e2004-00316-5)

Grenfell, B. T., Bjørnstad, O. N. & Finkenstädt, B. 2002 Dynamics of measles epidemics: scaling noise, determinism and predictability with the TSIR model. *Ecol. Monogr.* **72**, 185–202.

Halloran, M. E., Longini, I. M., Nizam, A. & Yang, Y. 2002 Containing bioterrorist smallpox. *Science* **298**, 1428–1432. (doi:10.1126/science.1074674)

Hatzimanikatis, V., Li, C., Ionita, J. A. & Broadbelt, L. J. 2004 Metabolic networks: enzyme function and metabolite structure. *Curr. Opin. Struct. Biol.* **14**, 300–306. (doi:10.1016/j.sbi.2004.04.004)

Heathcote, C. R. & Nicholls, D. F. 1990 Least-squares estimation of the contact rate in models for the spread of infectious diseases. *Biometrika* **77**, 161–168. (doi:10.1093/biomet/77.1.161)

Hethcote, H. W. & Yorke, J. A. 1984 *Gonorrhea transmission dynamics and control*. Lecture notes in biomathematics, vol. 56. Berlin, Germany: Springer.

Hochberg, M. E. 1991 Non-linear transmission rates and the dynamics of infectious disease. *J. Theor. Biol.* **153**, 301–321. (doi:10.1016/S0022-5193(05)80572-7)

Keeling, M. J. 1999 The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. B* **266**, 859–867. (doi:10.1098/rspb.1999.0716)

Keeling, M. J. 2005 The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.* **67**, 1–8. (doi:10.1016/j.tpb.2004.08.002)

Keeling, M. J. & Eames, K. T. D. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)

Keeling, M. J. & Grenfell, B. T. 2000 Individual-based perspectives on R_0. *J. Theor. Biol.* **203**, 51–61. (doi:10.1006/jtbi.1999.1064)

Kermack, W. O. & McKendrick, A. G. 1991 Contributions to the mathematical theory of epidemics I—1927. *Bull. Math. Biol.* **53**, 33–55.

Klovdahl, A. S., Dhofier, Z., Oddy, G., O'Hara, J., Stoutjesdijk, S. & Whish, A. 1977 Social networks in an urban area: first Canberra study. *Aust. N.Z.J. Sociol.* **13**, 169–172. (doi:10.1177/144078337701300215)

Levin, S. A. & Durrett, R. 1996 From individuals to epidemics. *Phil. Trans. R. Soc. B* **351**, 1615–1621. (doi:10.1098/rstb.1996.0145)

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. 2001 The web of human sexual networks. *Nature* **411**, 907–908. (doi:10.1038/35082140)

Liu, W., Hethcote, H. W. & Levin, S. A. 1987 Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.* **25**, 359–380. (doi:10.1007/BF00277162)

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)

Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. & Gerstein, M. 2004 Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312. (doi:10.1038/nature02782)

May, R. M. & Lloyd, A. L. 2001 Infection dynamics on scale-free networks. *Phys. Rev. E* **64**, 066 112. (doi:10.1103/PhysRevE.64.066112)

Meyers, L. A., Newman, M. E. J., Martin, M. & Schrag, S. 2003 Applying network theory to epidemics: control measures for *Mycoplasma pneumoniae* outbreaks. *Emerg. Infect. Dis.* **9**, 204–210.

Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M. & Brunham, R. C. 2005 Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* **232**, 71–81. (doi:10.1016/j.jtbi.2004.07.026)

Meyers, L. A., Newman, M. E. J. & Pourbohloul, B. 2006 Predicting epidemics on directed contact networks. *J. Theor. Biol.* **240**, 400–418. (doi:10.1016/j.jtbi.2005.10.004)

Mollison, D., Isham, V. & Grenfell, B. 1993 Epidemics: models and data. *J. R. Stat. Soc. A* **157**, 115–149. (doi:10.2307/2983509)

Molloy, M. & Reed, B. 1995 A critical point for random graphs with a given degree sequence. *Random Struct. Algor.* **6**, 161.

Moore, C. & Newman, M. E. J. 2000 Epidemics and percolation in small-world networks. *Phys. Rev. E* **61**, 5678–5682. (doi:10.1103/PhysRevE.61.5678)

Moreno, Y., Pastor-Satorras, R. & Vespignani, A. 2002 Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529. (doi:10.1007/s10051-002-8996-y)

Newman, M. J. E. 2002 The spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016 128. (doi:10.1103/PhysRevE.66.016128)

Pastor-Satorras, R. & Vespignani, A. 2001 Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **63**, 066 117. (doi:10.1103/PhysRevE.63.066117)

Pastor-Satorras, R. & Vespignani, A. 2002 Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E* **65**, 035 108. (doi:10.1103/PhysRevE.65.035108)

Petermann, T. & Rios, P. D. L. 2004 The role of clustering and gridlike odering in epidemic spreading. *Phys. Rev. E* **69**, 066 116. (doi:10.1103/PhysRevE.69.066116)

Potterat, J. J., Phillips-Plummer, L., Muth, S. Q., Rothenberg, R. B., Woodhouse, D. E., Maldonado-Long, T. S., Zimmerman, H. P. & Muth, J. B. 2002 Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sex Transm. Infect.* **78**, 159–163. (doi:10.1136/sti.78.6.467)

Poutanen, S. M. *et al.* 2003 Identification of severe acute respiratory syndrome in Canada. *N. Engl. J. Med.* **348**, 1995. (doi:10.1056/NEJMoa030634)

Rothenberg, R. B., Sterk, C., Toomey, K. E., Potterat, J. J., Johnson, D., Schrader, M. & Hatch, S. 1998 Using social network and ethnographic tools to evaluate syphilis transmission. *Sex Transm. Dis.* **25**, 154–160. (doi:10.1097/00007435-199803000-00009)

Roy, M. & Pascual, M. 2006 On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecol. Complex.* **3**, 80–90. (doi:10.1016/j.ecocom.2005.09.001)

Severo, N. C. 1969 Generalizations of some stochastic epidemic models. *Math. Biosci.* **4**, 169–191. (doi:10.1016/0025-5564(69)90019-4)

Shen, Z., Ning, F., Zhou, W., He, X., Lin, C., Chin, D. P., Zhu, Z. & Schuchat, A. 2004 Superspreading SARS events Beijing, 2003. *Emerg. Infect. Dis.* **10**, 256–260.

Shirley, M. D. F. & Rushton, S. P. 2005 The impacts of network topology on disease spread. *Ecol. Complex.* **2**, 287–299. (doi:10.1016/j.ecocom.2005.04.005)

Stroud, P. D., Sydoriak, S. J., Riese, J. M., Smith, J. P., Mniszewski, S. M. & Romero, P. R. 2006 Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Math. Biosci.* **203**, 301–318. (doi:10.1016/j.mbs.2006.01.007)

Stumpf, M. P., Wiuf, C. & May, R. M. 2005 Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc. Natl Acad. Sci. USA* **102**, 4221–4224. (doi:10.1073/pnas.0501179102)

Volz, E. In press. SIR dynamics in random networks with heterogeneous connectivity. *J. Math. Biol.*

Volz, E. & Meyers, L. A. Submitted. SIR epidemics in dynamic contact networks.

Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of "small-world" networks. *Nature* **393**, 440–442. (doi:10.1038/30918)

Wuchty, S. 2001 Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* **18**, 1694–1702.

Zachary, W. 1977 An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473.