Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Population Biology 67 (2005) 1–8
Theoretical Population Biology
www.elsevier.com/locate/ytpbi

The implications of network structure for epidemic dynamics

Matt Keeling*

Department of Biological Sciences and Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

Received 6 February 2004

Abstract

It has long been realised that the standard assumptions of mass-action mixing are a crude approximation of the true mechanistic processes that govern the transmission of infection. In particular, many infections can be considered to be spread through a limited network of contacts. Yet, despite the underlying discrepancies, mass-action models continue to be used and provide a remarkably accurate description of epidemic behaviour. Here, the differences between mass-action and network-based models are investigated. This allows us to determine when mass-action models are a reliable tool, and suggest ways in which their behaviour should be refined.

Keywords: Disease spread; Transmission networks; Clustering; Spatial correlation; Approximation

1. Introduction

The standard mass-action mixing models that describe the dynamics of infection within a population of hosts have proved to be a robust and surprisingly accurate predictive tool (Anderson and May, 1992; Mollison et al., 1993; Grenfell and Dobson, 1995). However, despite this unprecedented success, there are clear mechanistic difficulties with the basic mass-action assumptions. Mass-action models implicitly assume that each infected individual has a small chance of spreading infection to every susceptible individual in the population. In contrast, most diseases spread through a network of contacts, such that infection has a much higher probability of spreading to a more limited set of susceptible contacts (Klovdahl et al., 1994; Ghani et al., 1997; Rothenberg et al., 1998; Keeling, 1999; Moore and Newman, 2000; Eames and Keeling, 2002). The differences between these two approaches are examined here using simple analytical arguments augmented by stochastic simulation.

We consider in some detail how the standard mass-action models could be modified to account for the effects of network structure. Recent years have seen great progress in the understanding of disease spread through a variety of networks, including random graphs (Diekmann et al., 1990; Barbour and Mollison, 1990; Andersson, 1998; Neal, 2003), small-worlds (Watts and Strogatz, 1998; Kleczkowski and Grenfell, 1999; Moore and Newman, 2000) and scale-free networks (May and Lloyd, 2001; Pastor-Satorras and Vespignani, 2001). The vast majority of this research has utilised 'bottom-up' mechanistic approaches, where the individual-level network structure and behaviour is used to generate an approximation for the spread of infection. This type of approach has been very productive and provided many new insights into disease transmission. Here a very different, phenomenological approach is taken, where modifications to the standard ODE models of Kermack and McKendrick (1927) are sought that can capture the temporal dynamics of network based epidemics (Filipe and Maule, 2003). Hence, the modifications must capture both the local spatial correlations that rapidly develop in the early stages of the epidemic, as well as the larger scale spatial patterns that can influence the latter

*Fax: +44 24 765 246 19.
E-mail address: m.j.keeling@warwick.ac.uk.

0040-5809/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2004.08.002

dynamics. In contrast to the mechanistic approach, these phenomenological modifications can be more readily appreciated by epidemiologists who are familiar with the Kermack and McKendrick ODE models, although the form assumed for these modifications is far more speculative. The ultimate aim is to identify a modelling framework that can be readily parameterised from the type of case-reporting data that is available in the early phase of an epidemic, and that reproduces the type of behaviour observed in network simulations. Consequently, attention is given to how this model would be realised in the case of an observed outbreak and the implications for its control (Ferguson et al., 2003).

2. Initial variability

Let us consider the dynamics of an epidemic during the first generation, in particular the mean and the distribution of secondary cases from a primary infection. We compare the standard mass-action model to a regular random network where each individual has the same number of contacts and there is no clustering of connections (Andersson, 1998; Diekmann et al., 1998; Keeling, 1999; Neal, 2003). We suppose that the population size is large, therefore the number of secondary cases produced by the mass-action model over a given time period is approximately Poisson distributed. In contrast, for the network-based model the number of contacts per individual, n, is relatively small and therefore the distribution is Binomial (Keeling and Grenfell, 2000; Newman, 2001; Neal, 2003). If we equate the two modelling approaches, such that the expected number of secondary cases generated by the primary infection (R_0^1) is the same, then the mass-action assumption shows greater variance:

$$\mathrm{Var}_{\text{mass-action}} = R_0^1,$$

$$\mathrm{Var}_{\text{network}} = R_0^1\left(1-\frac{R_0^1}{n}\right),$$

$$\Rightarrow \ \mathrm{Var}_{\text{mass-action}} > \mathrm{Var}_{\text{network}}. \tag{1}$$

This difference increases as R_0^1 approaches the number of network contacts (n) and hence transmission between contacts becomes more certain. Similarly, the mass-action assumption also shows a far greater probability of extinction during the first generation, when the initial case fails to cause any secondary infections before recovering.

$$\mathrm{Prob}_{\text{mass-action}}(\text{extinction}) = \exp(-R_0^1),$$

$$\mathrm{Prob}_{\text{network}}(\text{extinction}) = \left(1-\frac{R_0^1}{n}\right)^n,$$

$$\Rightarrow \ \mathrm{Prob}_{\text{mass-action}}(\text{extinction}) > \mathrm{Prob}_{\text{network}}(\text{extinction}). \tag{2}$$

In this scenario, the discrepancy is greatest when both R_0^1 and n are relatively small (Fig. 1a)—although for biological reasons we are only interested in the case where R_0^1 is greater than one and the epidemic can propagate. As expected, when the number of contacts, n, becomes large the mass-action and network models give equivalent results.

While the assumption of equal number of secondary cases is reasonable for a theoretical comparison, it is more biologically plausible (although more mathematically involved) to compare models with the same basic rate of increase. From an applied perspective we are forcing both models to fit the same epidemic growth curve. Whereas the previous results hold true for any distribution of infectious periods, for simplicity we now

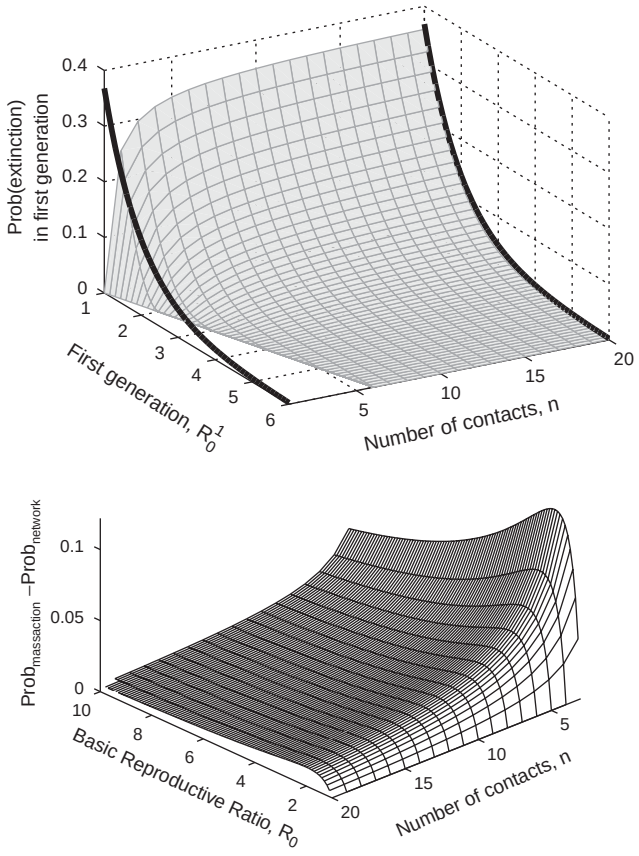

Fig. 1. Theoretical comparison between mass-action models and regular random network models in terms of the extinction probability. Graph A, the probability of extinction in the first generation for a random network with exactly n contacts per individual and random connections (grey mesh) can be seen to decrease with increasing R_0^1 but increase with increasing number of neighbours. The mass-action results (black lines) provide an asymptotic limit as n becomes large. Graph B, compares the probability of extinction before a large epidemic occurs; the discrepancy is greatest when there are few contacts and the asymptotic R_0 is close to two.

focus on Markovian models where the infectious period is exponentially distributed, although the methodology can be adapted to other distributions. Detailed modelling of connected pairs within the network shows that when there is no clustering the initial growth rate asymptotes to $(n-2)\tau - g$ (Keeling, 1999; Eames and Keeling, 2002); where τ is the rate of transmission across a connection and g is the recovery rate. In fact it is this early asymptotic growth rate that defines the true basic reproductive ratio of any structured disease model (Diekmann et al., 1990). We equate this to the rate of increase in the mass-action model, $(R_0-1)g$, by setting $\tau = R_0/(n-2)$. Using this equivalence we calculate the probability of extinction before a major epidemic ensues. For the mass-action model this generates the well-known result (Kermack and McKendrick, 1927) that:

$$\text{Prob}_{\text{mass-action}}(\text{extinction}) = \frac{1}{R_0}. \tag{3}$$

In contrast, the calculation for the network model is more involved. Once in the asymptotic exponential-growth phase the probability of extinction, P, is given by the solution to:

$$\begin{aligned} P &= \int_0^\infty g\,\exp(-gt)\sum_{m=0}^{n-1}\binom{n-1}{m}(1-\exp(-\tau t))^m \\ &\quad\times(\exp(-\tau t))^{n-m-1}P^m\,\mathrm{d}t \\ &= \int_0^\infty g\,\exp(-gt)[1+(1-\exp(-\tau t))(P-1)]^{n-1}\,\mathrm{d}t \\ &= \sum_{m=0}^{n-1}\binom{n-1}{m}P^{n-m-1}(1-P)^m\,\frac{n-2}{mR_0+n-2}, \end{aligned}$$

where the integral (t) is over all infectious period durations, the sum (m) is over $n-1$ contacts as one contact must have been the source of the infection, and $1-\exp(-\tau t)$ is the probability of transmission across a contact. However, the true probability of extinction must also take into account the fact that in the first generation there are n contacts that can be infected; therefore:

$$\begin{aligned} \text{Prob}_{\text{network}}(\text{extinction}) &= \sum_{m=0}^{n}\binom{n}{m}P^{n-m}(1-P)^m \\ &\quad\times\frac{n-2}{mR_0+n-2}. \end{aligned} \tag{4}$$

This probability is a specific case of the result by Newman (2001), and although this quantity cannot be calculated analytically, it can be easily solved numerically. Fig. 1b shows the difference between the two extinction probabilities for the mass-action and network assumptions; the network model has a much lower probability of extinction, with the greatest difference occurring when the number of contacts is low and R_0 is close to two. Interestingly, this region of parameter space matches that associated with many sexually transmitted diseases; therefore, although this result cannot replace the role played by high-risk core-groups (Hethcote and Vanark, 1987; Diekmann et al., 1990), it does suggest that STDs will persist better in reality than predicted by standard mass-action models.

Although the study of random-graphs with no clustering provides some analytical understanding of the differences between mean-field and network-based epidemics, the vast majority of epidemiological networks are highly clustered (Potterat et al., 1999; Liljeros et al., 2001; Girvan and Newman, 2002; Potterat et al., 2002; Newman and Park, 2003). The full dynamics on such complex networks are best captured using full stochastic simulation. We shall now consider how the relationships described above are modified by strong clustering of the network connections. Such clustering, defined by the presence of many triangular connections within the network, is a common feature of the type of social networks associated with the spread of airborne infection (Watts and Strogatz, 1998; Keeling, 1999; Andersson, 1999; Newman and Park, 2003; Read and Keeling, 2003).

3. Creation of clustered networks

Creation of the transmission network has four key components. The first step is to define a space in which individuals can be placed. For simplicity, and to aid visualisation, this is assumed to be a two-dimensional square (although higher-dimensional 'cubes' would also be feasible (Rhodes and Anderson, 1996)) which is scaled such that the average density of individuals is one per unit area. F focal points are now positioned at random within the space; such points are locations where individuals may congregate, and could represent schools and workplaces, or family groups, depending on their number. Individuals are now placed at a random location within the space; subsequently they are moved by a fixed proportion (f) towards their nearest focal point so as to increase the level of aggregation. Finally, individuals are connected to others as a function of the distance between them. Throughout a Gaussian function is used as the connection probability, whose height (H) and variance (V) can be varied to obtain the desired number of contacts and clustering within the network. Hence, two individuals separated by a distance d are connected with probability $H\exp(-d^2/(2V))$.

This methodology provides great flexibility in the types of networks that can be created, from randomised networks typical of sexually transmission networks (small f, large V) to highly clustered networks with very small cliques typical of airborne disease transmission predominantly within households (large F, large f, small V). Finally, for consistency between simulations we insist that the network must have a giant component, such that at least 90% of the population are connected

by a series of contacts; those networks that do not match this criterion are discarded. It is noted that sexual transmission networks are often very sparse and do not contain a giant component, however sexual networks are often far less clusters with fewer cliques making them more ameanable to more rigorous, mechanistic study (Eames and Keeling, 2002). Instead, attention is focused on the dynamics of airborne diseases, where the underlying network structure is much more complex and a giant component is present.

We could now investigate the disease dynamics on these networks in terms of the creation parameters (F, f, H and V); however, as our method of instantaneous network creation is very different from the complex and dynamic mechanisms involved in the true formation of contacts, these parameters have no clear biological interpretation and are unlikely to the measurable in practise. Instead we pick the creation parameters at random and measure three emergent properties of the network: the average number of contacts per individual ($\overline{n}$), the variance in the number of contacts (var(n)) and the clustering of the network (ϕ) defined as the ratio of triangular connections to all triple connections (Keeling, 1999). These should be measurable for a wide variety of social and sexual networks (Klovdahl et al., 1994; Potterat et al., 1999) and, we believe, provide a useful characterisation of the network (Rothenberg and Narramore, 1996).

For diseases on such networks we examine three basic aspects: the asymptotic growth rate, R_0, the probably of epidemic success and the best-fit mass-action model. These quantities are then related to the network properties $\overline{n}$, var(n) and ϕ.

3.1. Asymptotic R_0

As was noted for the simpler regular random network models, during the first few generations there is a reduction in the average number of cases produced by each infectious individual as the local density of susceptibles becomes rapidly depleted (Keeling, 1999). The asymptotic level that is soon reached defines the basic reproductive ratio, R_0, which is therefore dependent upon both the network structure and the disease characteristics. For the computer-generated networks two conflicting but interrelated factors influence R_0: the clustering and the heterogeneity of contacts. It has long been realised that heterogeneity in the number of contacts and in particular high-risk core groups increase the value of R_0 (Hethcote and Vanark, 1987; Diekmann et al., 1900; Anderson and May, 1992). When the network is highly clustered it is usually also highly assortative (Gupta et al., 1989); this will again cause further increases in R_0 as high-risk individuals are more likely to transmit to other high-risk individuals and hence increase the potential spread of infection (Gupta et al., 1989). However, clustering itself has a strong negative impact on R_0 as it reduces the number of new contacts in the second and subsequent generations (Keeling, 1999). This is because in a highly clustered network connected individuals share many common contacts and thus even in the early stages of an epidemic there is strong competition for susceptibles.

The effects of clustering on the growth-rate of infection are illustrated in Fig. 2a; only major epidemics are considered, those that fail in the early stages are discounted. Despite the great variation between networks, it appears that the clustering, ϕ, provides a reasonable measure for how network structure reduces R_0. In all these simulations individuals have an average

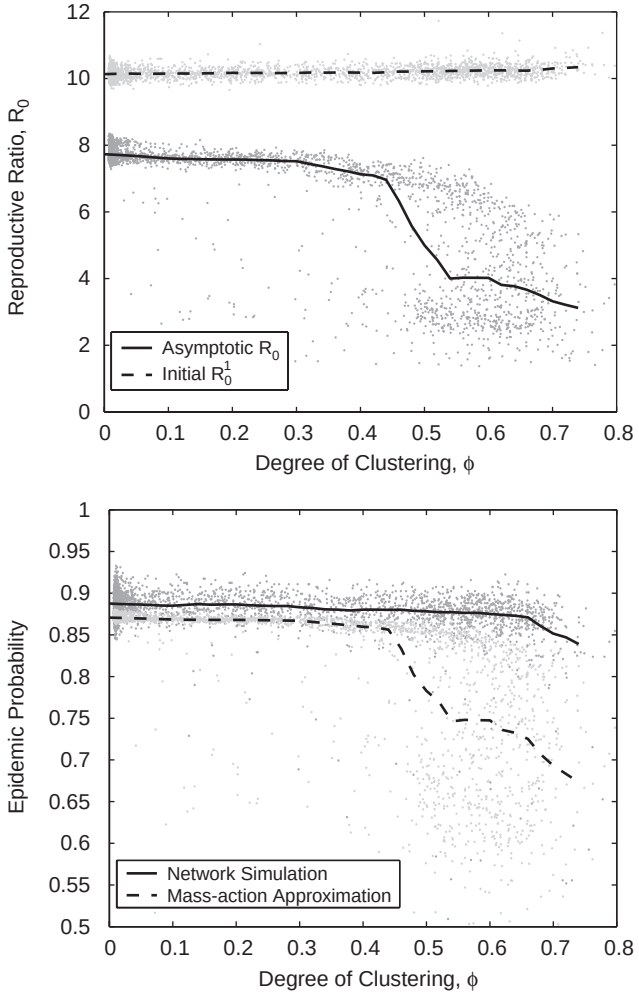

Fig. 2. Results from multiple simulations of a wide variety of networks constrained to have an average of 10 contacts per individual ($\overline{n} = 10$). The lines are calculated as the moving average of the median values. Graph A shows the basic reproductive ratio, R_0, both in terms of its true asymptotic value (dark grey, solid line) and the first generation R_0^1 (light grey, dashed line). Graph B gives the probability that a large epidemic is caused by a single seed infection. The results from the network simulations (dark grey, solid line) clearly lie above the theoretical values for mass-action models with the same asymptotic R_0 (light grey, dashed line).

of 10 contacts ($\overline{n} = 10$). Unclustered networks show that the asymptotic growth rate reduces to $R_0 \approx 8$ in agreement with the theory given above (Keeling, 1999), whereas when the clustering is maximal the basic reproductive ratio can be reduced as low as $R_0 \approx 2$. This reduction has a profound impact upon parameter estimation, and therefore the use of control measures, in the early stages of an epidemic. If control measures are implemented early (as is optimal) then it may be difficult to determine how much of the reduction in growth rate is due to the control and how much is due to the natural action of clustering. Ignoring the effects of clustering would lead us to over-estimate the success of any controls applied. In addition, there is a slight increase in the initial (first generation) R_0^1 with the degree of clustering due to the rise in contact heterogeneity and the fact that those epidemics that succeed are more likely to have started with highly connected individuals.

We are now in a position to test the likely success of an invading disease in a clustered network. Fig. 2b shows the probability that a single initial infection triggers a major epidemic (defined as at least one-third of the population) in the stochastic network model. Also shown is the theoretical probability of causing a major outbreak for a mass-action model with the same asymptotic R_0 value (Kermack and McKendrick, 1927). Clearly epidemics in clustered networks have a higher probability of epidemic success than their mass-action counterparts with equivalent R_0. This is for two reasons. Firstly, as was shown analytically, the binomial distribution associated with transmission to a limited number of contacts has less variability than the Poisson distribution associated with mass-action model. Secondly, the network model has a high average growth rate for the first generation or so (determined by R_0^1), thus increasing the persistence when the epidemic is most vulnerable to extinction.

3.2. Fitting a mass-action model

Information which can be used to parameterise disease transmission networks is extremely rare. There are a few studies that have attempted to map the entire network of sexual partnerships (Klovdahl et al., 1994; Wylie and Jolly, 2001), however these only provide information about a few isolated sub-populations and only inform about the spread of sexually transmitted diseases. It is likely that airborne disease have a very different network structure (Read and Keeling, 2003); some attempts have been made to assess these social networks using diary-based studies (Edmunds et al., 1997) but these are generally concerned with pair-wise contact levels and rarely give information about clustering.

In the absence of reliable data about the underlying network structure, epidemiologists are generally faced with the challenge of fitting mass-action models to case-report data (Mollison et al., 1993; Grenfell and Dobson, 1995). Following this approach, deterministic mass-action models are fitted to the average epidemic profile from stochastic network-based simulations. Three distinct parameters are varied to obtain the least-squares best-fit: the start-date of the epidemic, the transmission rate of the disease β_0, and the size of the initial susceptible population, S_0; it is assumed that the infectious period of the disease is known and that the distribution is exponential. Fig. 3 shows an example of the average epidemic from the network model and the associated deterministic mass-action results. Clearly, for the bulk of the epidemic there is an excellent agreement

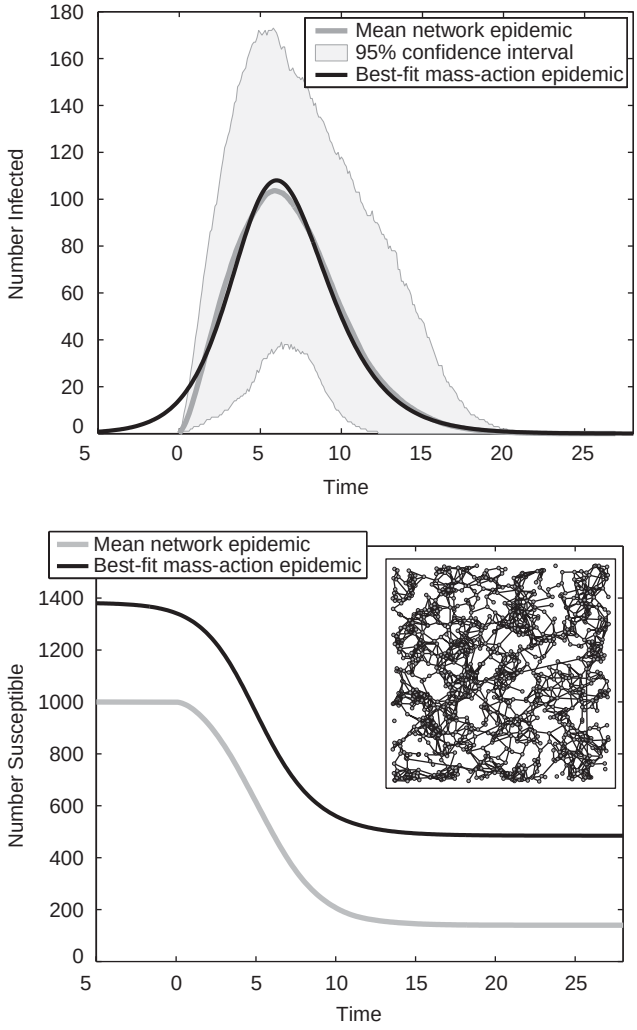

Fig. 3. For a given network of 1000 individuals (defined by $F = 0$, $H = 1$, $V = 1$, which leads to $\overline{n} = 6.106$, $\phi = 0.3652$) the average epidemic profile and the corresponding best-fit mass-action model ($\beta = 168424$, $S_0 = 1382$). Graph A shows the number of infected individuals over time, for the network model; the average of 1000 simulations is shown as well as the 95% confidence intervals. Graph B gives the associated susceptible population size; the inset graph gives an example of the type of network created. (Transmission rate across a network contact, $\tau = 1$, recovery rate $g = 1$.).

between the average simulation and the fitted model; this supports the continued use of simple mass-action models in many cases. However, there are two clear discrepancies. Firstly, the network-based model has a much faster initial growth rate (Fig. 3a), as detailed above. Thus if the mass-action model is our paradigm, it may often appear that during the early stages of the epidemic many cases have not been successfully identified. Secondly, the fitted mass-action model requires a higher initial level of the susceptibles, S_0 and leaves a higher level of susceptibles once the epidemic is complete (Fig. 3b).

We now extend the fitting procedure described above to a wide range of networks; Fig. 4a and b show the relative transmission rate, $\beta_0/(\overline{n}\tau)$, and the initial level of susceptibles, S_0, for the best-fit mass-action model. Again clustering can be seen to play a governing role in the dynamics. Initially, for a network with an average of $\overline{n}$ contacts per individual and an infection rate of τ across a contact, the initial transmission rate should be $\overline{n}\tau$. By comparing the fitted transmission rate, β_0, to this value, we observe the reduction due to network effects. This result is quantitatively similar to the reduction in R_0 observed in Fig. 2a, even though the mass-action model is derived by fitting to the entire epidemic rather than just the early exponential rise; again we find that highly clustered networks lead to a substantial decrease in the transmission of infection.

The scaling of the initial susceptible population, S_0 is less well defined; in Fig. 4b we consider the relative population size for clarity. The fact that network-based models exploit susceptibles with different efficiency compared to mass-action models is due to the spatial structure and the correlations that develop between susceptible and infectious individuals. In unclustered random networks, recovered individuals may completely block the transmission of infection; therefore the level of available susceptibles is much lower than the true level and hence the best-fit model will produce an under-estimate. In clustered networks however, aggregated clumps of susceptibles still exist in the latter stages of an epidemic and, once found, these can be rapidly infected as the local density of susceptibles is far higher than the global density; in such cases the best-fit model will over-estimate the level of susceptibles.

3.3. Approximating network-based models

The general patterns discussed above can be utilised to improve the predictive ability of the standard mass-action model. We shall assume, as is the norm, that we are attempting to parameterise a disease model from epidemic data. As shown above, while a successful fit can be established, both the level of susceptibles and the initial growth of the epidemic are subject to error. The following set of equations therefore offers a more

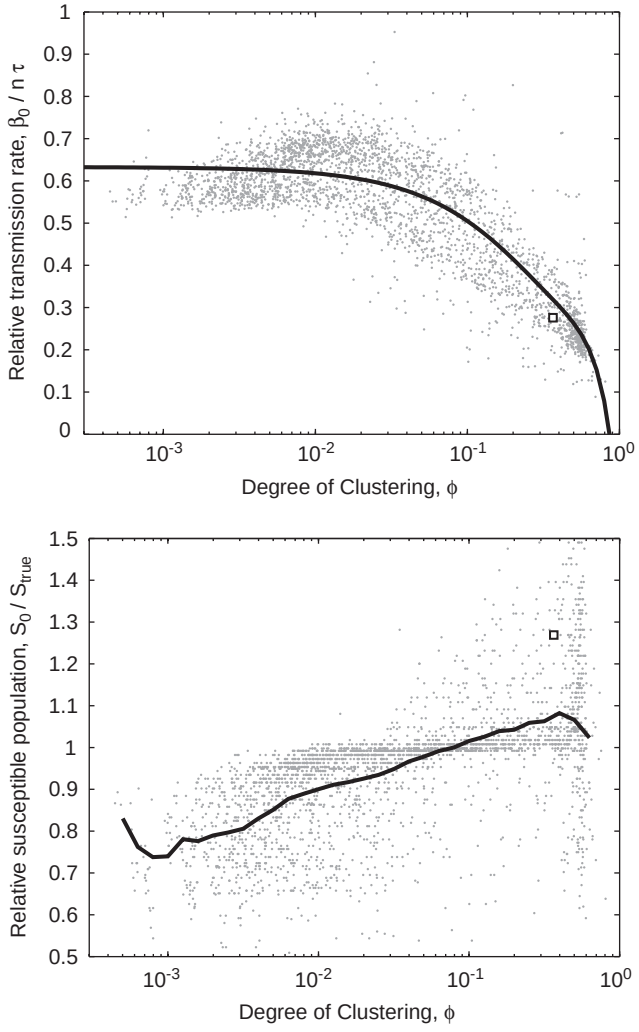

Fig. 4. For a range of networks that include a giant connected component (defining parameters $N = 1000$, $F \in [0, 1000]$, $f \in [0, 1]$, $V \in [10^{-4}, 10^{4}]$, $H \in [1, 10]/(\pi V)$) the two fitted parameters β_0 and S_0 from the associated mass-action model. For both graphs the relative value of the fitted parameter is shown, such that a value of one corresponds to the large $\overline{n}$ limit. The solid lines give a smoothed version of the local median values; the square indicates the parameters from Fig. 3.

accurate approximation:

$$\frac{d\widehat{S}}{dt} = -\beta(t)\widehat{S}\widehat{I}/\widehat{N},$$

$$\frac{d\widehat{I}}{dt} = \beta(t)\widehat{S}\widehat{I}/\widehat{N} - g\widehat{I},$$

$$\frac{d\widehat{R}}{dt} = g\widehat{I},$$

$$\beta(t) = \beta_0 + \beta_1 \exp(-gt),$$

$$S_{\text{true}} \approx \widehat{S} - S_1, \tag{5}$$

where β_0 and $\widehat{S}(0) = S_0$ are estimated from the epidemic profile and the parameters β_1 and S_1 depend on the

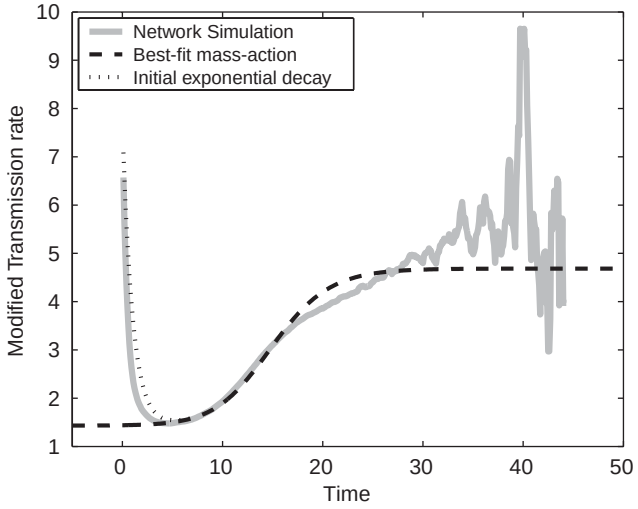

Fig. 5. For the three scenarios, the modified transmission rate defined as the per capita transmission rate, βS, divided by the proportion of the true population that are susceptible, S_{True}/N. The network model (grey line) shows some stochastic effects during the latter stages of the simulation when many epidemics are extinct. Modifying the level of susceptibles (dashed line, per capita transmission rate $=\beta_0\widehat{S}$) fits well to the mid and latter stages of the epidemic, whereas the inclusion of a exponential modifier (dotted line, per capita transmission rate $=\beta(t)\widehat{S}$) can also mimic the early behaviour. ($N=10^4$, $F=0$, $H=1$, $V=1$).

properties of the network, in particular the average number of neighbours and the degree of clustering. In general we would expect $\beta(0)=\beta_0+\beta_1=\overline{n}\tau$, such that spatial correlations and clustering do not effect the first generation dynamics.

The success of this refinement is shown in Fig. 5, which plots the modified transmission rate; that is the per capita transmission rate (which is $\beta\widehat{S}/\widehat{N}$ in the modified model) divided by the true proportion of susceptibles ($S_{\mathrm{true}}/N_{\mathrm{true}}$). This quantity defines how the per capita transmission is affected by spatial correlations as the epidemic progresses. If the standard mass-action model were accurate, this modified transmission rate would be constant and equal to $\overline{n}\tau\approx 6.1$. For the network simulation (grey line) the modified transmission rate is calculated as the average of $\tau[\mathrm{SI}]N/(\mathrm{SI})$, where [SI] refers to the number of susceptible-infected connected pairs (Keeling, 1999). The best-fit mass-action model (dashed-line) plots $\beta_0\widehat{S}N_{\mathrm{true}}/(S_{\mathrm{true}}\widehat{N})$, whereas the dotted line shows $\beta(t)\widehat{S}N_{\mathrm{true}}/(S_{\mathrm{true}}\widehat{N})$. Clearly the modification to β and the change in the susceptible population lead to very good agreement between the simple ODE model and the average network simulations.

4. Conclusion

It is clear that the structure of a contact network can have a profound effect on the dynamics of infectious diseases that are transmitted through it. Both a small average number of contacts and a high degree of clustering reduce the asymptotic basic reproductive ratio, R_0, compared to the initial growth rate. This can have profound effects on the estimation of parameters during the early stages of the epidemic. Of greatest concern is that such a reduction can swamp the effects of control measures and lead to an over-estimation of their effectiveness. Surprisingly, such network models are less prone to stochastic extinction than mass-action models with comparable dynamics. This is for two reasons, firstly the network model effectively has higher levels of transmission during the first generation (to compensate for the later decrease) when it is most vulnerable to extinction, and secondly the distribution of secondary cases in a network model is binomial which has a lower risk of extinction compared to the Poisson distribution of the mass-action models. This result has profound implications for the long-range spread of disease; due to the larger initial transmission rates with network-based transmission it may be far easier for an infection to invade and persist within a new population than its endemic (or even epidemic) dynamics suggest. All these results were found to hold across a variety of network structures, with the degree of clustering, ϕ, principally determining the differences between network-based and mass-action dynamics.

We have shown that to some degree the dynamics of diseases on networks can be captured with reasonable accuracy by simple modifications to the standard SIR equations. These modifications account for the initial reduction in the effective transmission rate over the first generations and the differences in efficient access to susceptibles due to the spatial structure that develops in the latter part of the epidemic. Again the scale of these modifications is primarily determined by the clustering of network contacts, with the most clustered networks having the greatest initial drop in transmission but a more efficient use of susceptibles later (Fig. 4). While such modifications help to improve the fit of the standard SIR model, even the un-modified model shows remarkable agreement with the more complex network-based simulations. This clearly supports the continued use of these traditional models in a wide variety of scenarios. The reason for this surprising success is that generally the dynamic spatial correlations, which distinguish network models, rapidly asymptote to quasi-equilibrium values; hence, the effective transmission rate and efficiency are usually set by network, are global parameters and show little variation throughout the epidemic with the effects of local fluctuations being rapidly dissipated.

In this paper, we have only considered the average results from multiple stochastic simulations where we have perfect knowledge of the epidemiological dynamics. The more practical application of assessing parameters during the course of a single noisy epidemic

with limited reporting accuracy is obviously more of a statistical challenge, although the same basic approach should still be valid.

Throughout we have used a plausible mechanism for constructing computer-generated networks, however this mechanism must instill certain properties and relationships on the network parameters. Many other construction mechanisms such as scale-free (May and Lloyd, 2001) or small-world (Moore and Newman, 2000) networks could be tested, and while preliminary investigations suggest that the same qualitative trends are observed there will be some quantitative differences. We have consistently found that the clustering of network contacts, ϕ, is one of the key parameters determining the dynamics of epidemics, however it is also one of the most difficult to measure for real interaction networks and has therefore received little attention. Given the applied importance of such characteristics, the estimation of clustering and other emergent network properties is a priority area for future research.

Acknowledgments

This research was funded by The Royal Society, BBSRC and The Wellcome Trust. I would also like to thank Bryan Grenfell, Jon Read and Ken Eames for their helpful comments.

References

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

Andersson, H., 1999. Epidemic models and social networks. Math. Scientist 24, 128–147.

Anderson, R.M., May, R.M., 1992. Infectious Diseases of Humans. Oxford University Press, Oxford.

Barbour, A., Mollison, D., 1990. Epidemics and random graphs. In: Gabriel, J.P., Lefevre, C., Picard, P. (Eds.), Stochastic Processes in Epidemic Theory. Springer, New York, pp. 86–89.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and the computation of the basic reproduction ratio R_0, in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28, 365–382.

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. Probab. 35, 448–462.

Eames, K.T.D., Keeling, M.J., 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proc. Natl. Acad. Sci. 99, 13330–13335.

Edmunds, W.J., O'Callaghan, C.J., Nokes, D.J., 1997. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc. Roy. Soc. London B 264, 949–957.

Ferguson, N.M., Keeling, M.J., Edmunds, W.J., Gani, R., Grenfell, B.T., Anderson, R.M., Leach, S., 2003. Planning for smallpox outbreaks. Nature 425, 681–685.

Filipe, J.A.N., Maule, M.M., 2003. Analytical methods for predicting the behaviour of population models with general spatial interactions. Math. Biosci. 183, 15–35.

Ghani, A.C., Swinton, J., Garnett, G.P., 1997. The role of sexual partnership networks in the epidemiology of gonorrhea. Sex. Transm. Dis. 24, 45–56.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99, 7821–7826.

Grenfell, B.T., Dobson, A., 1995. Ecology of Infectious Diseases in Natural Populations. Cambridge University Press, Cambridge.

Gupta, S., Anderson, R.M., May, R.M., 1989. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 3, 807–817.

Hethcote, H.W., Vanark, J.W., 1987. Epidemiological models for heterogeneous populations—proportionate mixing, parameter-estimation, and immunization programs. Math. Biosci. 84, 85–118.

Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. Roy. Soc. London B 266, 859–869.

Keeling, M.J., Grenfell, B.T., 2000. Individual-based perspectives on R_0. J. Theor. Biol. 203, 51–61.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. London A 115, 700–721.

Kleczkowski, A., Grenfell, B.T., 1999. Mean-field-type equations for spread of epidemics: the 'small-world' model. Physica A 274, 355–360.

Klovdahl, A.S., Potterat, J.J., Woodhouse, D.E., Muth, J.B., Muth, S.Q., Darrow, W.W., 1994. Social networks and infectious-disease—the colorado-springs study. Soc. Sci. Med. 38, 79–88.

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

May, R.M., Lloyd, A.L., 2001. Infection dynamics on scale-free networks. Phys. Rev. E 64, art. no. 066112.

Mollison, D., Isham, V., Grenfell, B., 1993. Epidemics: models and data. J. Roy. Statist. Soc. A 157, 115–149.

Moore, C., Newman, M.E.J., 2000. Epidemics and percolation in small-world networks. Phys. Rev. E 61, 5678–5682.

Neal, P., 2003. SIR epidemics on a Bernoulli random graph. J. Appl. Probab. 40, 779–782.

Newman, M.E.J., 2001. Spread of epidemic disease on networks. Phys. Rev. E 66, art. 016128.

Newman, M.E.J., Park, J., 2003. Why social networks are different from other types of network. Phys. Rev. E. 68, art. 036122.

Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86, 3200–3203.

Potterat, J.J., Rothenberg, R.B., Muth, S.Q., 1999. Network structural dynamics and infectious disease propagation. Int. J. STD. AIDS 10, 182–185.

Read, J.M., Keeling, M.J., 2003. Disease evolution on networks: the role of contact structure. Proc. Roy. Soc. London B 270, 699–708.

Rhodes, C.J., Anderson, R.M., 1996. Persistence and dynamics in lattice models of epidemic spread. J. Theor. Biol. 180, 125–133.

Rothenberg, R., Narramore, J., 1996. The relevance of social network concepts to sexually transmitted disease control. Sex. Transm. Dis. 23, 24–29.

Rothenberg, R.B., Potterat, J.J., Woodhouse, D.E., Muth, S.Q., Darrow, W.W., Klovdahl, A.S., 1998. Social network dynamics and HIV transmission. AIDS 12, 1529–1536.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Wylie, J.L., Jolly, A., 2001. Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. Sex. Transm. Dis. 28, 14–24.