

Parameterization of individual-based models: Comparisons with deterministic mean-field models

Darren M. Green*, Istvan Z. Kiss, Rowland R. Kao

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received 30 November 2004; received in revised form 27 July 2005; accepted 27 July 2005

Available online 8 September 2005

Abstract

The relating of deterministic, mean-field models into network models, where epidemic spread occurs between interconnected susceptible and infectious individuals or populations, requires careful consideration. Here, we discuss models that consider differently the manner in which contact rate and infectiousness change over time, with different algorithms suitable for different underlying processes. Though these models give coincidental results to the mean-field in the case of large, highly connected networks, the results when sparsely connected networks are considered may differ. Different subsets of the parameters from the mean-field epidemic (R_0 , generation time, infectiousness, etc.) are preserved in each case. Despite these differences, simulated epidemics generated under some model architectures are insensitive to the average degree of contact amongst nodes, k . Model-based estimates of k may be model dependent, and must therefore be viewed with caution.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Contact structure; Basic reproductive ratio; Transmission

1. Introduction

Disease dynamics are often analysed using compartmental models, and based on the solution of systems of ordinary differential equations where homogeneous mixing between different classes (i.e. the mean-field approximation) is assumed (Anderson and May, 1991). While these models have in many cases been highly successful, the mean-field assumption can be inappropriate when applied to certain systems, where interaction only occurs between a restricted subset of individuals. *In extremis* the case exists where the degree of contact is polarized: zero for some pairs of individuals, and non-zero for the rest. In this case, contact between individuals can be represented by a network of connections between interconnected ‘nodes’ (Albert and Barabási, 2002). This generic approach can represent a

variety of systems, such as individuals within a network of social interaction (Gupta et al., 1989; Kretzschmar and Wiessing, 1998; Newman, 2002; Meyers et al., 2005), or in a metapopulation-type system, populations of individuals (e.g. Ferguson et al., 2001; Hufnagel et al., 2005). Though both network and metapopulation models have existed for a long while, models combining the two approaches are less common.

R_0 , the basic reproductive ratio, is typically taken to mean

the average number of secondary infections produced when one infected individual is introduced into a [homogeneously mixed,] wholly susceptible host population at equilibrium (Anderson and May, 1991).

It is easily shown that for many biologically reasonable systems, $R_0 = 1$ is a global stability threshold for the disease-free state, a result generalized by Diekmann et al. (1990). In certain cases, R_0 can be estimated explicitly from the data, for example using field data

*Corresponding author. Tel.: +44 1865 281092.

E-mail address: darren.green@zoo.ox.ac.uk (D.M. Green).

from the recent 2001 FMD outbreak in the UK (Haydon et al., 2003). However, R_0 is normally indirectly estimated, for example by fitting model parameters to epidemiological data. Parameter estimates so derived are necessarily sensitive to the assumptions made elsewhere in the model; translating these parameter estimates to models of different architecture constructed with different assumptions can result in profound differences in model output. In many cases, the heterogeneity in transmission rates is critical for control (Roberts and Heesterbeek, 2003). R_0 is not necessarily the only parameter of interest here: estimates of the nature of the contact network structure, e.g. the mean number of connections per node k , are also likely to be model dependent, and k is of interest when considering the potential efficacy of contact tracing and disease control strategies (Kiss et al., 2005).

When moving from a differential equation model of epidemic spread to individual-based models, additional assumptions or a priori knowledge must be included concerning the spread of the epidemic at the level of the node; these may differ according to the nature of the node (for example, metapopulation or individual) and the nature of the disease. Below, we discuss a number of plausible, closely related network models of epidemic spread, incorporating different assumptions as to the mechanism of contact and the nature of the infectious period. We get the expected result of equivalency to the mean-field case as the network becomes completely connected, however markedly different results can be obtained when simulated on sparsely connected networks, with different properties preserved under different model assumptions. Thus, when considering the effect of degree of network connection on epidemic spread, it is necessary to consider carefully the design and purpose of the model.

2. Theoretical background

We consider four different sets of model assumptions, and compare them using the mean-field assumption as a convenient benchmark. While we do not imply that the mean-field is “better”, it is certainly the most common assumption; we shall discuss later the question of model appropriateness.

2.1. The mean field differential equation model

The starting point for the network models of epidemic spread considered below is the standard susceptible (S), infected/infectious (I) and removed (R) or SIR model, (discussed in detail elsewhere; see for example, Anderson

and May, 1991):

$$\begin{aligned}\dot{S} &= -\beta SI/N, \\ \dot{I} &= \beta SI/N - gI, \\ \dot{R} &= gI.\end{aligned}\tag{1}$$

Here, S , I , and R denote numbers of individuals, g the rate of conversion from I to R , and β is the number of potentially infectious contacts made per individual per unit time. Frequency-dependent contact is assumed, as this is a more appropriate starting point for translation to individual-based models than density dependence. In this model, R_0 is equal to β/g , and the instantaneous force of infection is $\lambda(t) = \beta(I/N)$.

Upon migration to an individual-based modelling approach, it is necessary to consider risk of infection across connections between pairs of individuals. We consider epidemic spread on undirected networks of connections between nodes, with a mean number of connections per node, k . The mean-field case is equivalent to an infinite number of nodes, N , with all possible connections between nodes available, i.e. $k = N - 1$.

2.2. The impact of transmissibility and network structure on epidemic spread

In network models, not only the properties of the individual, but also the nature of the network of connections between them is important in determining R_0 and the course of an epidemic (Keeling, 1999). It is well known that for sexually transmitted diseases, highly active individuals are extremely important. Anderson and May (1991) show that

$$R_0 \propto (1 + C_V^2),\tag{2}$$

where C_V is the coefficient of variation of the distribution of the number of contacts per individual. This result is due to the correlations between susceptibility and infectiousness. Should these be uncorrelated, the expression reverts to the intuitive understanding of the second paragraph above. We note that in undirected networks these are always correlated. May and Lloyd (2001), further define the “transmission potential”, ρ_0 , such that $\rho_0 = \beta/g$ with parameters as described for the mean field model introduced above, and

$$R_0 = \rho_0(1 + C_V^2).\tag{3}$$

They recognized that for the mean-field model $\rho_0 = R_0$. Newman (2002) defines the probability of infection occurring across a connection between an infected and susceptible node through the whole infectious period as the “transmissibility”, T ; this is equivalent to \bar{q} as defined by Diekmann and Heesterbeek (2000). As with ρ_0 , this is a property of the node and its connections, not the structure of the network as a

whole, and the two are related such that $\rho_0 = kT$. Diekmann and Heesterbeek (2000) note that in undirected networks, the susceptible neighbourhood of an infected node is reduced as it is connected to its source infection. In this case R_0 is given by

$$R_0 = \bar{q}(k - 1 + \sigma^2/k),$$

therefore

$$R_0 = \rho_0(1 + C_V^2 - 1/k), \tag{4}$$

where σ^2 is the variance of the number of connections per node. Keeling (1999) similarly found that R_0 in network models is reduced compared with that in mean-field models due to the nature of the correlation between susceptible and infectious nodes.

Below, we concentrate on the effect of model parameterization upon the “transmission potential”, ρ_0 : the number of secondary cases generated from index cases chosen at random from the population, which is intimately related to T (Newman, 2002) or \bar{q} (Diekmann and Heesterbeek, 2000). Though for the mean field model this is equal to R_0 , this is not generally the case for network models, as shown above. However, both these quantities are sometimes (erroneously) referred to as R_0 . We retain the usage ρ_0 throughout, except where explicitly discussing the mean-field model itself.

2.3. Model Ia—rate of infectious contact independent of k

The naïve assumption when constructing a network epidemic model equivalent to a mean-field model with known transmission potential ρ_0^* ($= R_0$), is that β should be kept constant across different networks, implying that each infectious individual is capable of infecting a known number of susceptible individuals per unit time. Thus, for networks where each node has a higher number of connections to other nodes, the probability of each connection resulting in transmission goes down. Given this, the rate of infectious contact across each connection, τ , should be defined as

$$\tau = \frac{\beta}{k} = \frac{g\rho_0^*}{k}, \tag{5}$$

i.e. simply dividing the potentially infectious contacts amongst the connected individuals. In the mean-field case, k is equal to $N - 1$, and there is no local exhaustion of susceptible nodes around an index case during its infectious period. Thus in this limit the expectation value of ρ_0 is that of the standard SIR model:

$$\rho_0 = \frac{1}{g} k\tau \approx \frac{\beta}{g} (= \rho_0^*). \tag{6}$$

However, Keeling (1999) notes that estimates of ρ_0 derived from mean field models do not apply in the case

of a network model with small k . Following Keeling and Grenfell (2000), consider a single isolated infectious node (the index case) in a network of otherwise susceptible nodes with an average of k connections per node. In a random network, the distribution of the numbers of connections will be binomial, well approximated by a Poisson distribution for large N and small k . The probability that the infected node remains infectious at time t after the beginning of the infectious period is given by an exponential function $H(t)$ with a corresponding probability density function (PDF) of removal at time t in $(0, \infty)$ $h(t)$:

$$\begin{aligned} h(t) &= g \exp(-gt), \\ H(t) &= \exp(-gt). \end{aligned} \tag{7}$$

Similarly, the probability that a susceptible node in contact with a single infected node remains susceptible at time t after the beginning of the infectious period is given by an exponential function $S(t)$, which has a corresponding PDF of infection at time t , $s(t)$:

$$\begin{aligned} s(t) &= \tau \exp(-\tau t), \\ S(t) &= \exp(-\tau t). \end{aligned} \tag{8}$$

For a node with exactly m connections, the expected number of secondary infections at infection age u is given by

$$R(m, u) = m(1 - S(u)). \tag{9}$$

Summation over all possible infectious period lengths and numbers of connections, treating these as independent, gives the expectation of the number of secondary from a single infected node ρ_0 :

$$\rho_0 = \sum_{m=0}^{\infty} \text{Poisson}(m, k) \int_{u=0}^{\infty} h(u)R(m, u) du. \tag{10}$$

This can be shown to reduce to

$$\rho_0 = \frac{\tau}{\tau + g} \sum_{m=0}^{\infty} m \text{Poisson}(m, k), \tag{11}$$

$$\rho_0 = k \frac{\tau}{\tau + g}, \tag{12}$$

as given for exponentially distributed infectious periods given by Keeling and Grenfell (2000) (cf. Diekmann et al., 1998). The summation term of Eq. (11) evaluates to the expectation of whichever distribution is used. Therefore, the same equation, Eq. (12), is obtained not only for a Poisson distribution of m , but for any degree distribution with a finite k , for example with constant m or in finite scale-free networks (Albert and Barabási, 2002). For networks with a low heterogeneity in k , the basic reproduction ratio R_0 will approach ρ_0 .

2.4. Model IB— ρ_0 independent of k

Parameterization according to Model Ia allows β to be kept constant across networks, but at the expense of preserving ρ_0^* . Where ρ_0^* is assumed to be known, for example when estimated directly from contact tracing data (Haydon et al., 2003), the result of Keeling and Grenfell (2000) can be used to infer the β required to obtain the known ρ_0^* . Rearrangement of Eq. (12) above gives the following definition of τ such that ρ_0^* is preserved across networks with different k :

$$\tau = g \frac{\rho_0^*}{k - \rho_0^*}, \quad k > \rho_0^*. \tag{13}$$

2.5. Model IIa—constant rate of generation of new cases

In Models Ia and Ib, as susceptibles connected to an infectious node become infected, there are fewer of them to infect; since infectiousness is constant, the average number of secondary cases created per unit time for a given infectious node therefore decays exponentially. This results in the higher incidence found in Model Ib for networks with higher k (Fig. 1). To fit the network model to the incidence curve found for the mean-field model, infectiousness must therefore rise over time, should k be small.

We model this case as follows: For an index case with exactly m connections, the number of secondary infections in an infectious period of length u is given by

$$R(m, u) = \int_{t=0}^u (m - R) C_{IS}(t) \psi(t) dt, \tag{14}$$

where C_{IS} denotes the contact rate per susceptible neighbour, and ψ infectiousness at time t after infection. If ψC_{IS} is inversely proportional to the number of susceptible neighbours $m - R$, then the above reduces to

$R(m, u) = \alpha u$, independent of m , where the constant α is the instantaneous rate of production of new infections per infected node. The expectation value of ρ_0 is thus

$$\rho_0 = \sum_{m=0}^{\infty} \text{Poisson}(m, k) \int_{u=0}^{\infty} \alpha u h(u) du = \frac{\alpha}{g}. \tag{15}$$

Therefore, α is equal to $g\rho_0$, and corresponds to β in Model I above. For a network with an average of k connections per node, the probability of any connection causing infection during the whole infectious period of length u is therefore

$$P = u \frac{g\rho_0^*}{k}. \tag{16}$$

For large u and small k , this function can return probabilities greater than unity. In this case, a value of 1.0 is assumed and all connections are considered as causing infection, and the model output deviates from the mean-field result.

2.6. Model IIb—fixed-length infectious periods

For some epidemics, a constant infectious period would be more applicable than the exponential distribution considered in Models I and IIa, which for most epidemics will overestimate the number of nodes with infectious period considerably less than the mean (Lloyd, 2001). Therefore, in a variant of Model IIa, IIb, we assume constant $u = 2/g$ for all nodes, rather than the exponential distribution obtained from Eq. (7). This ensures that the mean age of infection when secondary infection occurs, $1/g$, agrees with that for the earlier models (Section 4.2). However, with an infectious period twice as long as for the exponential distribution, the prevalence of infection is doubled. Prevalence for Model IIb is therefore plotted divided by two to allow for fairer comparison.

2.7. The next generation

Except for the index case, in the undirected networks discussed here, all subsequent infected nodes with m undirected connections have at most $m-1$ possible infectious connections, as one connection leads back to its source infection (Diekmann and Heesterbeek, 2000). Where m is distributed according to a Poisson distribution, as in the simulations described below, then R_0 as calculated from Eq. (3) equals ρ_0 and epidemics are similar across different values for k . This is a special case, and this result will not be recovered for some other distributions of m . For the calculation of R_0 in a network context, not only the properties of the nodes and connections (i.e. the transmissibility) but also the network structure must be taken into account. In this paper, we consider specifically model parameterization for constant ρ_0 .

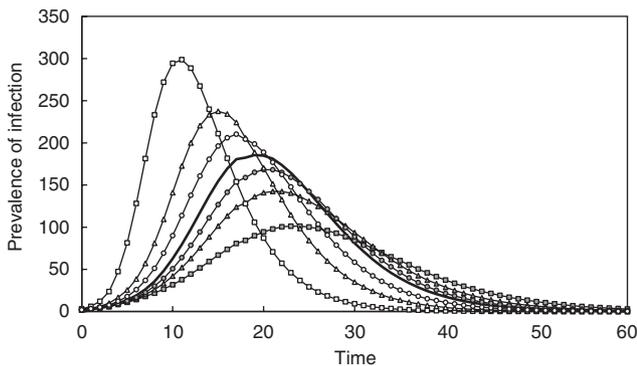


Fig. 1. Epidemic time courses for Models Ia and Ib: number of infectious individuals, averaged over 1000 epidemic and network realizations for $N = 1000$, $\gamma = 1$ and $1/g = 4$. Shaded symbols: τ determined by Eq. (5) (Model Ib) and $\rho_0^* = 2.5$; open symbols: τ determined by Eq. (13) (Model Ia). ■ $k = 5$; ▲ $k = 10$; ● $k = 20$. The mean field equivalent model is shown with no symbol.

3. Method

3.1. Network construction

A number $N = 1000$ nodes were uniformly, randomly distributed over a square landscape at positions x_i , the same realization of which was used throughout, with the two coordinates of each node chosen independently from a uniform distribution. The mean density of nodes γ was set at one node per square unit of the landscape. Random, well-mixed networks, without clustering, were constructed between these nodes, where the probability of connection between any two nodes i and j is given by

$$\pi_{ij} = k/(N - 1), \quad i \neq j. \quad (17)$$

An undirected connection between nodes i and j exists if and only if $U(0, 1) < \pi_{ij}$, where $U(a, b)$ is defined as a function that returns a random number sampled uniformly from the interval $[a, b]$. As connections are undirected, and there are no self-loops, only the cases for all $i < j$ need to be examined.

Random networks as generated above do not account for the clustering that is likely to be present in many real networks, where certain groups of individuals are more likely to be connected to each other than to members of other such groups. Here clustered networks were generated by allowing for a greater chance of connection between nodes that are located close together in space, using the equation

$$\pi_{ij} = \frac{K}{2\pi D^2} \exp\left(\frac{-d_{ij}^2}{2D^2}\right), \quad i \neq j, \quad (18)$$

$$\pi_{ij} = 0, \quad i = j,$$

where D adjusts the average length of a connection, K the number of connections per node, and $d_{ij} = |x_i - x_j|$ is the Cartesian distance between nodes. For very close pairs of nodes, Eq. (18) results in probabilities of connection greater than unity. In this case, a probability of 1.0 is assumed and the pair of nodes are connected.

On a finite network, Eq. (18) does not result in a network with number of connections per node K and average connection length D (Read and Keeling, 2003). Therefore, the average number of connections per node, k , and the average connection length l were measured on the generated networks, and stepwise adjustment of both parameters K and D performed to obtain the k and l to within a desired tolerance; these parameters are shown in Table 1. Table 1 also indicates the clustering coefficient of the networks generated, ϕ . This is defined as the probability that if connections exist between a node A and two others, B and C , then B and C themselves will be connected.

Table 1

Required parameters K and D in Eqs. (17) and (18) for desired k and l in constructed networks for $N = 1000$, $\gamma = 1$

k	l	K	D	ϕ
<i>Clustered networks</i>				
5	2	5.41 (0.019)	1.62 (0.005)	0.11
10		10.7 (0.029)	1.63 (0.004)	0.23
20		23.1 (0.12)	1.56 (0.003)	0.47
<i>Unclassified networks</i>				
5	16	5	—	0.005
10	(Measured)	10	—	0.01
20		20	—	0.02

Standard errors of the parameter estimates are shown in brackets ($n = 10$). Clustering coefficients ϕ for the generated networks are shown.

3.2. Epidemic simulation

An epidemic was initiated by seeding the network with a single index case, and then simulated using asynchronous updating. In the first two models (Models Ia and Ib), the time elapsed between infection and removal for each node was determined as an exponential deviate:

$$u = -\ln(U(0, 1))/g. \quad (19)$$

The time elapsed between infection of a node and secondary infection of a connected node q was calculated in a similar manner to u , though such infection was conditional on the infected node remaining infectious after this interval, and the connected node not having already been infected. The probability of infection across all connections is considered a constant. Thus, no interaction between τ and k is assumed:

$$q = -\ln(U(0, 1))/\tau. \quad (20)$$

In the third model (Model IIa), secondary infectious cases were uniformly distributed through the infectious period of a node with duration u (defined as in Eq. (19)), subject to a probability of infection of each connected node of τu . For the limit of large N and large k ($N, k \rightarrow \infty$; $N = k - 1$), mean field behaviour is obtained from models I and IIa, with $\beta = \rho_0^* g = \tau k$.

Comparison simulations were carried out as in Model I, but under the mean-field condition that encounters between nodes are entirely random. This was implemented by constructing a network where each node is connected to every other node, and thus k was a constant $N-1$ for all nodes. This is a stochastic form of the differential equation model shown in Eq. (1), similar results to which are obtained with a large population N and large initial seeding.

4. Results

4.1. Behaviour of Model I

Time courses of epidemics parameterized according to Models Ia and Ib are contrasted in Fig. 1 for unclustered networks and various values for k . Alongside these is shown the time course for the mean-field case ($k = N - 1$). With increasing k , the variance of the number of secondary cases derived from a primary case increases (though the expectation ρ_0 is constant in Model Ib). This results in a greater probability of stochastic extinction during the initial phase of the epidemic for higher k . To account for this, Fig. 1 considers only ‘established’ epidemics, with final epidemic size $Z > 200$.

Results from Models Ia and Ib were most similar to each other and to the mean-field case for large k ($k = 20$). For Model Ia, higher k corresponds to an earlier, higher epidemic peak and a larger final epidemic size Z —as mentioned earlier, ρ_0 is not preserved across different values of k in Model Ia, and thus, neither is the final epidemic size. The opposite effect is seen for Model Ib: as expected ρ_0 values are the same, but the epidemic peaks higher and earlier for lower k . The final epidemic size here was independent of k .

Observed ρ_0 was measured on the simulated epidemics by determining the average number of infected cases having the index case as their source of infection. This average was consistently found to be slightly lower than that predicted from Eq. (12). This results from the index cases competing for susceptible nodes with its daughter infections. The magnitude of this effect was greater in clustered networks where the number of completely connected triplets of nodes is high (Keeling, 1999). Since ρ_0 here is, as mentioned earlier, equivalent to May and Lloyd’s (2001) ‘force of infection’, it is not affected by the form of the degree distribution.

4.2. Epidemic generation time

The shape of the prevalence curves shown in Fig. 1 and elsewhere incorporate not only ρ_0 , but also on the generation time T_{II} . Should ρ_0 and T_{II} vary with both k and model architecture, then it is possible that with knowledge of some of the above, some inference could be made about the others. Thus, below, we examine the relationship between k and T_{II} for the model architectures described earlier.

For Model Ib, measured generation time across the whole epidemic, defined as the average period between the time of an infection and the time of infection of its source node, increased monotonically with time in both clustered and unclustered networks though was similar initially (Fig. 2). This results from an increasing ratio of R to I nodes through time, resulting in fewer encounters between I and S nodes and lower infection rates.

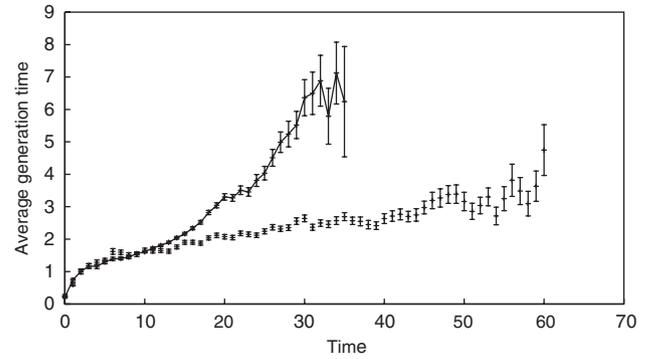


Fig. 2. Mean generation time: Model Ib for clustered and unclustered networks. Parameters as in Table 1 and Fig. 1 with $k = 10$. Means and their standard errors are shown for nodes becoming infected at time t averaged over 50 network and epidemic realizations. Results for clustered networks (unconnected symbols) up to $t = 60$ and unclustered networks (connected symbols) up to $t = 35$ are shown.

The curve for clustered networks in Fig. 2 lies to the right of that for unclustered networks as the former produces epidemics that take a much larger number of generations to spread over the whole landscape. Averaged across every infection event of a whole epidemic however, no difference in generation time was found between clustered and unclustered networks, nor for simulations with different N ($N = 500$ and 1500 were tested). Nevertheless, generation time was, as mentioned above, dependent on k with shorter generation times for lower k (Fig. 3).

The behaviour of epidemic generation time can be studied qualitatively by considering the time to infection in isolated pairs of I–S nodes. The PDF describing the probability of secondary infection in an I–S pair at time t after the I node becomes infectious is given by

$$j(t) = H(t)s(t). \quad (21)$$

The average time of an infection in the first generation of infection, after the infection of the index case, denoted by T_{II0} , is thus given by the expectation of this function across all t , conditional on the probability that infection of the S node occurs, $\tau/(\tau + g)$:

$$T_{II0} = \int_{t=0}^{\infty} \frac{tj(t)}{\tau/(\tau + g)} dt, \quad (22)$$

$$T_{II0} = \frac{1}{\tau + g}, \quad (23)$$

where τ is parameterized according to Models Ia and Ib, T_{II0} varies with k in both cases:

$$T_{II0} = \frac{1}{g} \left(\frac{k}{k + \rho_0^*} \right) \quad \tau = \frac{g\rho_0^*}{k} \quad (\text{Model Ia}), \quad (24)$$

$$T_{II0} = \frac{1}{g} \left(\frac{k - \rho_0^*}{k} \right) \quad \tau = g \frac{\rho_0^*}{k - \rho_0^*} \quad (\text{Model Ib}). \quad (25)$$

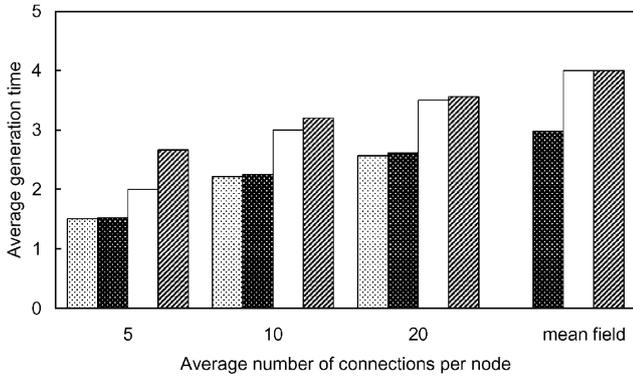


Fig. 3. Generation time for Model Ib simulations with different values of k , compared with T_{II0} as predicted from Eq. (25) together with the mean-field approximation. Parameters as in Table 1 and Fig. 1. The average generation time for all infections over 2500 network and epidemic realizations are shown. Light-shaded bars: clustered network; dark-shaded bars: random network; white bars: T_{II0} from Eq. (25); and striped bars: T_{II0} from Eq. (24).

The solution of Eqs. (24) and (25) for different values of k are shown for comparison with the measured generation times in Fig. 3. T_{II0} is consistently higher than measured generation time, and higher for larger k , but this effect is smaller for T_{II0} in Model Ia, as defined by Eq. (24).

4.3. Epidemics with a constant time course and different k

Fig. 4 shows simulation results for Model IIa. For larger k , Model IIa agrees with the mean-field approximation shown also in Fig. 1, and has the same ρ_0 , generation time, final epidemic size, and epidemic time course. For small k ($k=5$), some nodes with u considerably greater than $1/g$ will have insufficient connections to produce the large number of secondary cases predicted for large u , and for the definition of Eq. (16), allows for the non-sensical result $p > 1$. As a result, for smaller k , epidemics with a smaller ρ_0 are obtained.

For large N and k , the mean time to infection of Model IIa, T_{II0} , can be shown to equal $1/g$, which is the same value found for the mean-field approximation. Dividing Eq. (16) by u gives a daily rate of infection per connection τ whose definition is identical to that shown in Eq. (5). Thus, for large N and k , model behaviour approaches once more that of the mean-field approximation, and the simple definition of Eq. (5) applies:

$$\tau = \frac{P}{u} = \frac{g\rho_0^*}{k} \tag{26}$$

Where a constant infectious period is modelled, $p > 1$ as described above does not occur. Model IIb amends Model IIa above such that a constant infectious period of length $2/g$ is assumed for all nodes. Thus, $T_{II0} = 1/g$ as before. The time course of such an epidemic is shown

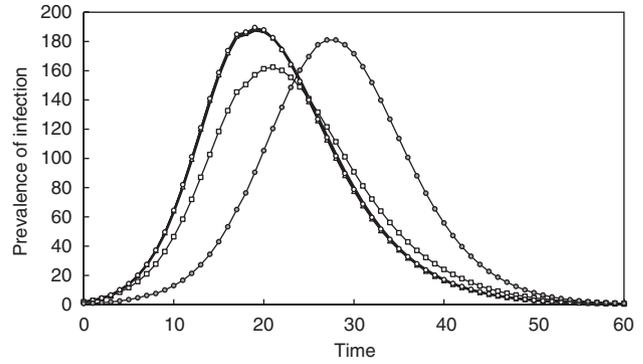


Fig. 4. Epidemic time courses for Model IIa: number of infectious individuals, averaged over 2500 epidemic and network realizations. Parameters as in Table 1 and Fig. 1. \square $k=5$; \triangle $k=10$; \circ $k=20$ (open symbols). The mean field equivalent model is shown as a solid line with no symbol. (Curves for $k=10$, $k=20$, and the mean-field equivalent coincide.) For comparison, the time course for Model IIb is also shown (\bullet , solid symbol), with prevalence adjusted for the longer mean infectious period.

Table 2
Variation in model outputs with varied k and constant β or ρ_0^* for the different models

Parameter	Model Ia	Model Ib	Model IIa	Model IIb
<i>Parameters independent of k</i>				
ρ_0	No	Yes	Yes ^a	Yes
β	Yes	No	Yes ^a	Yes
Z	No	Yes	Yes ^a	Yes
Generation time	No	No	Yes ^a	Yes
T_{II0}	No	No	Yes ^a	Yes
Final epidemic size	No	Yes	Yes ^a	Yes
Approaches mean-field for large k	Yes	Yes	Yes	No

^aFor $k \gg \rho_0^*$.

in Fig. 4. Here, the incidence of infection was divided by two to account for each individual being infectious for twice as long. Simulation results for this model were independent of k , with a constant epidemic time course, but did not coincide with the mean field approximation for large k .

4.4. Comparison of model behaviour

With the exception of Model IIb, all the models described above generated the same results when approximating mean field conditions with $k \approx N \rightarrow \infty$. For smaller k , the different models preserve different parameters from this epidemic time course (Table 2). In Model IIa, epidemic generation time, final epidemic size, and ρ_0 were independent of k where k is somewhat larger than ρ_0^* . For Model Ib, ρ_0 and final epidemic size did not vary with k , but epidemic generation time was lower for

lower k . Model Ia had only an infection rate β that was independent of k . Only in Model IIb were all parameters independent of k , but here, the mean field model was not approached for large k due to the different underlying assumption concerning infectious period distribution.

5. Discussion

Of the models described above, all but Model IIb use the rate parameter g to determine infectious period length according to an exponential distribution, which can have consequences for epidemic dynamics (Lloyd, 2001). Model IIb addresses this by using fixed infectious period lengths—the delta distribution—which for some infections will be more reasonable; more complex models have used forms of the gamma distribution (Lloyd, 2001), for example by using multiple sequential infectious stages (Ferguson et al., 2001).

Which of the above modelling approaches is the most appropriate will depend upon the system being studied. Model I assumes equal probability of contact between pairs of nodes in all states throughout the infectious period and constant infectiousness. This would be appropriate for a disease where the pattern of contact between individuals is unaffected by the presence of the infectious agent (for example, if there is a long delay before clinical symptoms are manifest), or where the infectiousness of contact is not altered by changes in pattern of contact (as might be the case for STDs).

For other systems, infectiousness over time, disease incidence and k may be interrelated. For example, in the recent epidemic of foot-and-mouth disease in the UK in 2001, the number of secondary cases arising from infected premises generated over time remained roughly constant (Haydon et al., 2003). This may have been because k was high and therefore there was little exhaustion of connected individuals, in which case it is consistent with the original ODE model (Eq. (1)). However, should k have been low, infectiousness may have been increasing. This is one interpretation of Model II: the probability of contact is unchanged from Model I, but infectiousness of nodes increases through the infectious period in such a manner that the force of infection is fixed.

On the other hand, patterns of contact may change. An alternative interpretation of Model II is that the force of infection is constant over time; this can be regarded as assuming that outward connections are not made to infected nodes. If there is a fixed total duration of contact, then this time would be divided amongst all available remaining nodes and thus the transmission rate to each susceptible node is increasing (e.g. transmission of the Black Death to susceptible villages by fleeing individuals). This constant rate of generation of secondary cases cannot however be

maintained indefinitely: for nodes with small k or long infectious periods, the supply of available connected nodes will be exhausted before the infectious period ends.

In epidemiology, one of the important uses of network models is in the analysis of the efficacy of contact tracing (Huerta and Tsimring, 2002; Eames and Keeling, 2003; Kiss et al., 2005; Meyers et al., 2005). One obvious consequence is that for large k , contact tracing is much less effective than for small k (Kiss et al., 2005). Therefore, when considering disease control policies with regards to contact tracing, a determination of k is essential. However, the results above show that sensitivity of model output to k is dependent on the exact model architecture used, and therefore whether a value for k can be meaningfully obtained from epidemic data depends upon using a suitable model to describe the system. For systems resembling Model I, given an epidemic time course and either of the generation time and R_0 , an estimate of k could be made by fitting the models. In other such systems, this may not be the case.

The results presented above show that small differences in model architecture, unimportant when the case of a large, fully connected network is considered, can produce large differences in model output when sparsely connected networks of infection are considered, but where these differences appear in model output depends on where these differences in the model architecture lie. In particular, ρ_0 , R_0 , and the epidemic generation time may or may not be preserved across different values of k , depending upon the particular modelling assumptions made. Thus, for such network models, parameter estimates derived from one such model may not translate safely to another. Additionally, where model output is insensitive to the degree of connection amongst nodes, no estimate of the degree of connection can be obtained from fitting the model to epidemic data, important if predictions of control are based on model dependent assumptions (e.g. Kao, 2003).

Acknowledgments

RRK and IZK are funded by the Wellcome Trust. DMG is funded by DEFRA.

References

- Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Modern Phys.* 74, 47–97.
- Anderson, R.M., May, R.M., 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
- Diekmann, O., Heesterbeek, J.A.P., 2000. *Mathematical Epidemiology of Infectious Diseases*. Wiley, Chichester, UK.

- Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and computation of the basic reproduction ratio R_0 in models of infectious diseases for 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., 2003. Contact tracing and disease control. *Proc. R. Soc. Lond. B* 270, 2565–2571.
- Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* 292, 1155–1160.
- 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.
- Haydon, D.T., Chase-Topping, M., Shaw, D.J., Matthews, L., Friar, J.K., Wilesmith, J., Woolhouse, M.E.J., 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. R. Soc. Lond. B* 270, 121–127.
- Huerta, R., Tsimring, L.S., 2002. Contact tracing and epidemics control in social networks. *Phys. Rev. E* 66, 056115.
- Hufnagel, L., Brockmann, D., Geisel, T., 2005. Forecast and control of epidemics in a globalized world. *PNAS* 101 (42), 15124–15129, doi:10.1073/pnas.0308344101.
- Kao, R.R., 2003. The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease. *Proc. R. Soc. Lond. B* 270, 2557–2564.
- Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* 266, 859–867.
- Keeling, M.J., Grenfell, B.T., 2000. Individual-based perspectives on R_0 . *J. Theor. Biol.* 203, 51–61.
- Kiss, I.Z., Green, D.M., Kao, R.R., 2005. Disease contact tracing in random and clustered networks. *Proc. R. Soc. B* 272, 1407–1414.
- Kretzschmar, M., Wiessing, L.G., 1998. Modelling the spread of HIV in social networks of injecting drug users. *AIDS* 12, 801–811.
- Lloyd, A.L., 2001. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor. Popul. Biol.* 60, 59–71.
- May, R.M., Lloyd, A.L., 2001. Infection dynamics on scale-free networks. *Phys. Rev. E* 64, 066112.
- Meyers, L.A., Pourbohloul, B., Newman, M.E.J., Skowronski, D.M., Brunham, R.C., 2005. Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* 232, 71–81.
- Newman, M.E.J., 2002. The spread of epidemic disease on networks. *Phys. Rev. E* 66, 016128.
- Read, J.M., Keeling, M.J., 2003. Disease evolution on networks: the role of contact structure. *Proc. R. Soc. Lond. B* 270, 699–708.
- Roberts, M.G., Heesterbeek, J.A.P., 2003. A new method for estimating the effort required to control an infectious disease. *Proc. R. Soc. Lond. B* 270, 1359–1364.