THE UNIVERSITY OF

WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with
publisher policies. Please scroll down to view the document
itself. Please refer to the repository record for this item and our
policy information available from the repository home page for
further information.

To see the final version of this paper please visit the publisher's website.
Access to the published version may require a subscription.

JOURNAL
OF
THE ROYAL
SOCIETY

Interface

Insights from unifying modern approximations to infections on networks

Thomas House and Matt J. Keeling

Supplementary data	"Data Supplement" http://rsif.royalsocietypublishing.org/content/suppl/2010/06/09/rsif.2010.0179.DC1.html
References	**This article cites 24 articles, 8 of which can be accessed free** http://rsif.royalsocietypublishing.org/content/early/2010/06/09/rsif.2010.0179.full.html#ref-list-1
P<P	Published online 10 June 2010 in advance of the print journal.
EXiS Open Choice	This article is free to access
Rapid response	Respond to this article http://rsif.royalsocietypublishing.org/letters/submit/royinterface;rsif.2010.0179v1
Subject collections	Articles on similar topics can be found in the following collections biomathematics (201 articles) computational biology (188 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Insights from unifying modern approximations to infections on networks

Thomas House* and Matt J. Keeling

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

Networks are increasingly central to modern science owing to their ability to conceptualize multiple interacting components of a complex system. As a specific example of this, understanding the implications of contact network structure for the transmission of infectious diseases remains a key issue in epidemiology. Three broad approaches to this problem exist: explicit simulation; derivation of exact results for special networks; and dynamical approximations. This paper focuses on the last of these approaches, and makes two main contributions. Firstly, formal mathematical links are demonstrated between several prima facie unrelated dynamical approximations. And secondly, these links are used to derive two novel dynamical models for network epidemiology, which are compared against explicit stochastic simulation. The success of these new models provides improved understanding about the interaction of network structure and transmission dynamics.

Keywords: epidemic; network; transmission; pairwise; simulation; infection

1. INTRODUCTION

The vast majority of infectious diseases can be considered as spread through a network of contacts between individuals or groups. These network concepts have been used, to great effect, in scenarios where the network of relevant contacts can be readily ascertained: for sexually transmitted infections, the network of human sexual contacts (Klovdahl 1985; Rothenberg *et al.* 1998; Jolly & Wylie 2002); and for bovine diseases, the network of animal movements as captured by the Cattle Tracing System (Kao *et al.* 2006; Lysons 2007; Vernon & Keeling 2009). More recently, attention has focused on the implications of network structure for human infections transmitted through close contact (such as influenza, SARS and smallpox) using diary-based social encounter information (Mossong *et al.* 2008), contact tracing (Ghani *et al.* 2009) or predicted patterns of movements (Eubank *et al.* 2004) to infer the appropriate network structure. In all cases, it is important to acknowledge the effects of network structure, both because of its impact on the uncontrolled epidemic (Keeling 1999; Volz & Meyers 2007) but also because of its utility in targeting controls and tracing the spread of infection (Eames 2006). It is, therefore, important that we develop methods of modelling infection dynamics on complex structured networks. Stochastic simulations obviously provide the most accurate and versatile models, but at the expense of tractability. There has, therefore, been considerable focus on developing less computationally intensive approaches that can help us interpret the difference between network-based predictions and those of mean-field (random mixing) models where local structure is ignored.

A recent review paper (Bansal *et al.* 2007) counter-posed three dynamical approaches to the study of epidemic dynamics on networks. These are: pairwise approaches (Keeling 1999), dynamic probability generating function (PGF) formalism (Volz & Meyers 2007; Volz 2008) and heterogeneous mixing (Dieckmann & Heesterbeek 2000; Moreno *et al.* 2002). Following this approach, we demonstrate here that each of these three approaches is an approximation to the more general pairwise model of Eames & Keeling (2002). We then use the relationships between approaches to derive two new moment-closure-based models for network epidemics: a 'clustered PGF' model, which is capable of capturing epidemic dynamics on clustered networks of heterogeneous link distribution using a relatively small number of ordinary differential equations (ODEs), and a 'heterogeneous susceptible–infectious–susceptible (SIS)' model, which makes a significant but less dramatic reduction of dimensionality where acquired immunity is not long-lasting. Finally, we compare these two models to simulation on exemplar networks similar to those considered in more applied contexts.

2. DERIVING OTHER APPROACHES FROM PAIRWISE MODELS

We start by considering a general heterogeneous contact network with N nodes. Using the notation

---

*Author for correspondence (t.a.house@warwick.ac.uk).

developed in Keeling (1999) and Eames & Keeling (2002), we use square brackets [] to represent the expected numbers of nodes, pairs or triples of any particular type; all notation used is summarized in table 1. The local structure of this network can be defined in terms of three main measures: the degree of distribution d_k (where d_k is the proportion of nodes having k network contacts), the clustering of contacts ϕ that measures the ratio of triangles (groups of three nodes all connected to each other) in the network to triples of all types (lines of three nodes with or without a transitive link) and the assortativity within the network as captured by the matrix

$$C_{k,l} = \frac{\bar{n}[kl]N}{k[k]l[l]},$$

where $\bar{n}$ is the mean node degree, which compares the true number of pairs of a given degree with the expected number if half-links connected at random. The matrix $\mathcal{C}$ represents the extent to which the number of $[kl]$ pairs is over- or under-represented with respect to a random process of pairing half-links (Newman 2003). We wish to explore how these three forms of local structure influence the types of model formulation that can be successfully applied.

The full pairwise equations of Eames & Keeling (2002), from which we begin our analysis, are given for SIR-type infections by

$$
\begin{aligned}
[\dot{S}_k] &= -\tau[S_kI], \\
[\dot{I}_k] &= \tau[S_kI] - \gamma[I_k], \\
[S_k\dot{S}_l] &= -\tau\sum_m([S_kS_lI_m] + [S_lS_kI_m]), \\
[S_k\dot{I}_l] &= \tau\left(\sum_m([S_kS_lI_m] - [I_mS_kI_l]) - [S_kI_l]\right) \\
&\quad - \gamma[S_kI_l], \\
[I_k\dot{I}_l] &= \tau\left(\sum_m([I_mS_kI_l] + [I_mS_lI_k])\right. \\
&\quad \left. + [S_kI_l] + [S_lI_k]\right) - 2\gamma[I_kI_l], \\
[S_k\dot{R}_l] &= -\tau\sum_m[I_mS_kR_l] + \gamma[S_kI_l], \\
\text{and}\quad [I_k\dot{R}_l] &= \tau\sum_m[I_mS_kR_l] = \gamma([I_kI_l] - [I_kR_l]),
\end{aligned}
$$
(2.1)

where $[A_k]$ refers to the number of nodes of type A with degree k; a disease state without a subscript implicitly contains the sum over all possible degrees (e.g. $[A_kB] = \sum_l[A_kB_l]$); τ is the transmission rate; and γ is the recovery rate. (The equations for infections that obey SIS-type dynamics are derived by modifying how recovery acts.) As with many moment-based methods, these equations are exact but unclosed. If we continue to write higher-order equations for triples, quads, etc., then these

equations will close only once we reach the network size N, destroying the main motivation for the use of ODEs to describe epidemics on networks. We, therefore, must seek approximations that will allow us to close these equations at a lower dimension (approximating the number of triples in terms of pairs and singles), which will aid computation and analytic understanding.

2.1. Dynamical and network assumptions

We present now a series of different assumptions about network structure and epidemiological dynamics that allow other approaches to be derived from this general pairwise model, discussing the reasoning behind each assumption and its range of validity. We pay particular attention to the number of equations required to simulate SIR dynamics in terms of the maximum node degree M.

2.1.1. Triple closure. Closure schemes for pairwise models, which have become essentially standard, have been presented for networks of significant clustering but with homogeneous degree (Keeling 1999) and also for unclustered networks with heterogeneous degree (Eames & Keeling 2002). The natural combination of these two schemes is

$$
\begin{aligned}
[A_kB_lC_m] \approx \frac{(l-1)}{l}\left((1-\phi)\frac{[A_kB_l][B_lC_m]}{[B_l]}\right. \\
\left. +\phi\frac{\bar{n}N}{km}\frac{[A_kB_l][B_lC_m][C_mA_k]}{[A_k][B_l][C_m]}\right),
\end{aligned}
$$
(2.2)

where A, B and C stand for arbitrary disease states. While this closure is the natural extension of existing approximations, it does pose a question of interpretation compared to existing closures, since the term proportional to ϕ cannot be rigorously interpreted as a prevalence of triangles owing to the lack of symmetry between the k, l and m nodes. An extended discussion of how to interpret asymmetric clustered closure appears in House & Keeling (2010). It is also worth noting that the unclustered pairwise equations can be reinterpreted in terms of neighbourhoods (Dangerfield *et al.* 2008), although this approach has yet to be extended to the clustered case.

The closure scheme (2.2) acting on the full pairwise equations (2.1) produces a closed ODE system of $(5M+2)M$ independent equations for SIR dynamics, which can quickly outstrip even modern computational resources for graphs with 'fat tailed' degree distributions such as scale-free networks, where M is often very large. This motivates the investigation of further approximations that can allow us to reduce the system size.

One additional, commonly made assumption that will be used in deriving other approaches is the absence of triangles in the network, that is,

$$\phi = 0. \tag{2.3}$$

<div style="text-align:center">Table 1. Notation.</div>

symbol	description
N	number of nodes in the network
M	maximum node degree
$[k]$	number of nodes of degree k (equal to $\sum_A [A_k]$)
$[kl]$	number of pairs with one member having degree k, and with the other having degree l (equal to $\sum_{A,\,B}[A_k B_l]$)
$\bar{n}$	average degree distribution (equal to $\sum_k k[k]/N$)
d_k	proportion of nodes with degree k (equal to $[k]/N$)
$\mathcal{C}_{kl}$	correlation matrix between nodes of degree k and degree l
ϕ	clustering coefficient of the network (equal to the number of triangles divided by the number of triples)
$[A_k]$	number of nodes in state A with k neighbours
$[A]$	number of nodes in state A (equal to $\sum_k [A_k]$)
$[A_k B_l]$	number of pairs with one member in state A and with degree k, and with the other member in state B and with degree l
$[A_k B]$	number of pairs with one member in state A and with degree k, and with the other member in state B (equal to $\sum_l [A_k B_l]$)
$[AB]$	number of pairs with one member in state A, and with the other member in state B (equal to $\sum_k [A_k B]$)
$[A_k B_l C_m]$	number of triples with one edge member in state A and with degree k, with the middle member in state B and with degree l, and with the other edge member in state C and with degree m
$\theta(t)$	the fraction of degree one nodes that remain susceptible at time t
$Y(t)$	auxiliary variable used in clustered PGF model (equal to $\sum_k k[I_k]$)
$g(x)$	PGF for the network degree distribution (equal to $\sum d_k x^k$)
τ	rate of transmission of infection across a network link
γ	rate of recovery from infection

2.1.2. Deconvolution of pairs. The assumption introduced by Eames & Keeling (2002) was to express the joint probability of a fully described pair as a product involving pair types and network structure:

$$[A_k B_l] \approx \frac{[A_k B][B_l A]}{[AB]} \frac{[kl]\bar{n}N}{k[k]l[l]}. \qquad (2.4)$$

Under this assumption, pairs with a single indexing degree are the main variables, which significantly reduces the number of independent ODEs needed in the pairwise equations to $10M$ for SIR dynamics. *A priori* the accuracy of this assumption is not assured, since it depends on the precise dynamical process taking place on the network; however, for all epidemiological scenarios so far considered in the literature, the results using this approximation are in extremely close quantitative agreement with the full pairwise model. Most importantly, this assumption allows the consideration of assortative mixing and heterogeneous degree, in a system where a number of differential equations is

linear rather than quadratic in the maximum node degree.

2.1.3. Detailed balance. The most general prevalence of pairs of connected nodes with degrees k and l is given by

$$[kl] = \mathcal{C}_{k,l} \frac{k[k]l[l]}{\bar{n}N},$$

that is, if $\mathcal{C}_{k,\,l}$ can take any value for combinations of k and l, then all possible distributions for $[kl]$ can be enumerated. However, an assumption is often made that the correlation matrix $\mathcal{C}$ obeys

$$\mathcal{C}_{k,l} = 1 \quad \forall k, l. \qquad (2.5)$$

In mean-field (non-network) models, such an assumption has been used, to great effect, to study the spread of sexually transmitted infections in risk-structured populations defined by individual-level data on sexual contacts (Anderson & May 1992); however, it is recognized that, in general, most populations are assortative such that $\mathcal{C}_{k,\,l} > 1$ for similar k, l and $\mathcal{C}_{k,l} < 1$ for dissimilar k, l. In network models, the assumption that $\mathcal{C}_{k,\,l} = 1$ means that link 'stubs' from each node are connected randomly, which is typically called *detailed balance* with respect to swapping randomly picked edges in physical science (Newman 2003), although terminology and definitions concerning assortativity can be different in other subject areas, for example, the sexually transmitted infection literature (e.g. Ghani & Garnett 1998).

2.1.4. Pair closure. If network dynamics are effectively dominated by assortative mixing of risk classes, then we may wish to remove pair-level variables through the assumption

$$[A_k B_l] \approx \frac{[kl]}{[k][l]} [A_k][B_l], \qquad (2.6)$$

which allows us to keep assortativity but loses the effects of network structure—the correlation between the states of connected nodes is lost—although some of these can be maintained through extra factors, as in Kiss *et al.* (2006). Application of this closure reduces the number of equations needed to $2M$ for SIR dynamics.

2.1.5. Deconvolution of individuals. The final approximation we consider has not been previously explicitly stated in this form, but can be used to derive the PGF approach from a general pairwise model. It can be viewed as the main assumption that allows the PGF formalism to approximate disease dynamics on heterogeneous networks within a low dimensional framework. This involves writing the joint probability $[A_k B]$ (defined above) as a product, and is analogous to the concepts used to generate equation (2.4), in that its validity depends on the independence of dynamics and network structure.

$$[A_k B] \approx [AB] \frac{k[A_k]}{\sum_l l[A_l]}. \qquad (2.7)$$

The underlying approximation is that given a node (of type A_k), the type of connected node (taken here to be B) is independent of k and therefore independent of the local network structure. For SIR dynamics, this assumption creates a set of equations whose dimension does not depend on the maximum node degree M, although the exact number of equations required depends on whether clustering is present, and which quantities one wishes to calculate over the course of an epidemic.

2.2. Relating ordinary differential equation-based approaches

Starting with the general pairwise model (2.1), the assumptions above can be used to derive other approaches. Taking the assumptions (2.2)–(2.5) and (2.7) allows us to derive the PGF equations originally formulated from first principles in Volz (2008) (full equations given in electronic supplementary material). This is an interesting and unexpected result since the underlying arguments used originally to derive the pairwise and PGF models are quite different. The standard heterogeneous-mixing models associated with risk-structured populations (given in electronic supplementary material) are derived by putting the assumption (2.6) into the pairwise equations and ignoring the pair-level variables.

We also note that for networks with no clustering and degenerate degree distribution (often called either regular graphs or homogeneous random networks), the PGF and pairwise approaches are formally identical, since assumption (2.7) is trivially satisfied.

3. CONSTRUCTION OF NOVEL MODELS

While the formal links between prima facie distinct epidemic models are intrinsically interesting, the main motivation for our work is to derive novel, parsimonious models for epidemics on a range of complex networks without having to argue the formulation from first principles. The fact that full pairwise models are simply written down, together with the new explicit form of the assumption underlying PGF models (2.7), means that it is possible to produce systematically much simpler ODE-based models for network-based epidemics. We now consider two such models, together with the potential limits on this methodology.

3.1. Incorporation of clustering

The relationships between different ODE approaches to network epidemics presented above open up the intriguing possibility of extending the PGF approach to include clustering so that both heterogeneity in link distribution and clustering can be analysed using a low-dimensional model with a small number of dynamical variables. This is indeed possible: using (2.4) and (2.7) together with the standard pairwise closure (2.2), the two triples that appear in the unclosed

pairwise SIR equations can be closed through

$$
\left.
\begin{aligned}
[SSI] &\approx [SS][SI]\frac{g''(\theta)}{N(g'(\theta))^2} \\
&\quad \left((1-\phi) + \phi\bar{n}\frac{[SI]}{\theta g'(\theta)\,Y}\right), \\
\text{and}\quad [ISI] &\approx [SI]^2\frac{g''(\theta)}{N(g'(\theta))^2}\left((1-\phi) + \phi\bar{n}N\frac{[II]}{Y^2}\right),
\end{aligned}
\right\}
\tag{3.1}
$$

where we define $\theta(t)$ as the fraction of degree 1 nodes that remain susceptible at time t and

$$
Y = \sum_k k[I_k],
\tag{3.2}
$$

and $g(x) = \sum_k d_k x^k$ is the PGF for the node degree distribution. This gives a new dynamical system that, together with the closure relations (3.1), determines the epidemic behaviour:

$$
\left.
\begin{aligned}
\dot{\theta} &= -\tau\frac{[SI]}{Ng'(\theta)}, \\
[\dot{I}] &= \tau[SI] - \gamma[I], \\
\dot{Y} &= \tau\left(\frac{\theta g''(\theta)}{g'(\theta)} + 1\right)[SI] - \gamma Y, \\
[\dot{SS}] &= -2\tau[SSI], \\
[\dot{SI}] &= \tau([SSI] - [ISI] - [SI]) - \gamma[SI], \\
\text{and}\quad [\dot{II}] &= 2\tau([ISI] + [SI]) - 2\gamma[II].
\end{aligned}
\right\}
\tag{3.3}
$$

The number of susceptibles can be tracked non-dynamically through the relationship

$$
[S] = Ng(\theta).
\tag{3.4}
$$

The primary significance of this result is that, regardless of the maximum node degree on a network, a system of six ODEs together with the non-dynamical relationship between $[S]$ and θ can be used to calculate an expected prevalence curve for heterogeneous, clustered networks.

3.2. Incorporation of other disease natural histories

In addition to the clustered PGF model derived above, which assumes SIR (susceptible–infectious–recovered) dynamics, other extensions of the PGF approach will clearly be possible, for example, the inclusion of a latent class or multiple infectious classes as in Kamp (2009). While this requires more variables, the system will still be of far lower dimension than the equivalent full pairwise model.

However, other commonly studied disease behaviours, such as the SIS paradigm, which is ideal for sexually transmitted infections and has been extensively studied using pairwise approaches, are not straightforwardly reduced to the PGF formulation. This is because, under SIS dynamics, the set of ODEs governing the evolution of quantities such as $\sum_k k^a$

$[A_k]$ is not closed. This does not, however, preclude the use of approximation (2.7). Inserting this assumption into the standard pairwise SIS model gives the following new model:

$$
\left.
\begin{aligned}
&[\dot{S}_k] = -\tau[SI]\frac{k[S_k]}{\sum_l l[S_l]} + \gamma([k] - [S_k]), \\
&\text{and} \quad [\dot{SI}] = \tau[SI]\left(\sum_k k[S_k] - 2[SI]\right)\frac{\sum_l l(l-1)[S_l]}{(\sum_m m[S_m])^2} \\
&\qquad - (\tau + \gamma)[SI] + \gamma\left(\sum_k([k] - [S_k]) - [SI]\right).
\end{aligned}
\right\}
$$
(3.5)

Unlike the clustered PGF model above, the number of these equations that need to be manipulated numerically is $M + 1$, that is, linear in the maximum node degree; however, their dimensionality is still significantly lower than standard pairwise models and, perhaps more importantly, there are only two equations that need to be manipulated in analytic work on this system.

3.3. Assortativity

Deviation from null assortativity is a complexity that simply cannot be naturally incorporated in the PGF equations without vastly increasing the dimensionality of the system. We provide a numerical demonstration of the importance of this observation in the electronic supplementary material. Nevertheless, as shown in Eames & Keeling (2002), by using networks (2.2) and (2.4), an assortative model can be created that is linear in maximum node degree.

3.4. Comparison with simulation

We now test each of the two new models above (clustered PGF and heterogeneous SIS) against simulation and compare to other relevant ODE approaches.

For the clustered PGF, we start with a random graph of approximately 10^4 nodes and with Poisson parameter $\lambda = 6$. We then introduce a clustering coefficient of $\phi = 0.2$ using the 'big V' rewiring (Bansal *et al.* 2009; House & Keeling 2010), which does not change the degree distribution. Following Serrano & Boguñá (2006), we bias node selection by $k(k - 1)$, so individual-level clustering is constant; while this will make C_{kl} slightly deviate from unity, at the level of clustering we consider this should have negligible dynamical impact. We choose these network parameters because they separate the ODE approaches without being so large that concerns about global network properties like giant component size are posed. While it is probable that realistic networks for respiratory infection have more variable degree distributions, larger numbers of mean nodes and higher clustering coefficients, our approach is to start with a system with few underlying parameters and a transparent method for the introduction of clustering. Epidemic rates were $\tau = 0.8$, $\gamma = 1$. Figure 1 shows the comparison of the clustered PGF model, which is in good agreement with simulation, and two other ODE approaches. The homogeneous
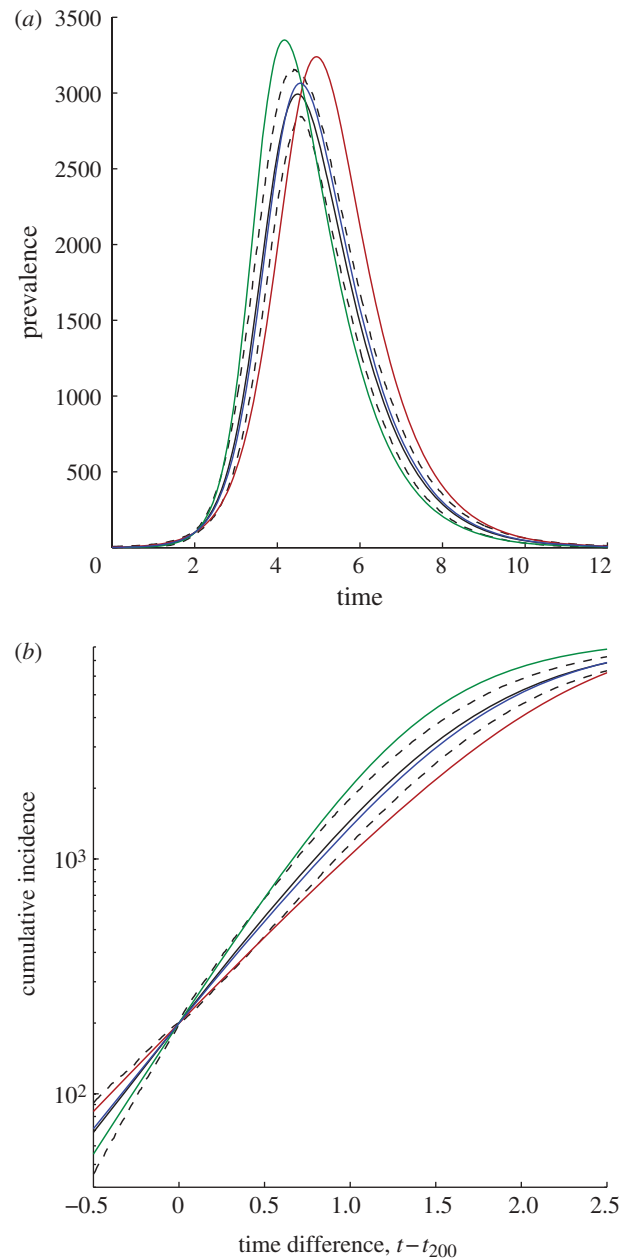

Figure 1. Numerical test of the clustered PGF model against simulation and other ODE approaches. The network has size $N \approx 10^4$, its degree of distribution is Poisson with mean $\bar{n} = 6$ and the clustering coefficient is $\phi = 0.2$. The transmission rate is $\tau = 0.8$ at unit recovery rate. We shift time for each of 10^3 stochastic simulations, so all curves agree on when a cumulative incidence of 200 is reached, and the simulation mean and prediction interval can be meaningfully visualized. Clearly, the clustered PGF approach is in excellent agreement with simulation. Solid line, simulation mean; dashed line, simulation 95% PI; red line, homogeneous pairwise; green line, PGF; blue line, clustered PGF.

pairwise model underestimates early growth in cases and, conversely, the unclustered PGF model overestimates early growth. This is as would be expected from general arguments about the effects of clustering (Keeling 1999) and population heterogeneity (Dieckmann & Heesterbeek 2000). The electronic supplementary material shows how these other, existing, ODE approaches remain in good agreement with

simulations on appropriate networks and also considers the impact of assortativity.

For the heterogeneous SIS model, we generate a scale-free network of approximately 10^4 nodes by using the standard Barabási–Albert (1999) method with parameters $m_0 = 20$, $m = 2$, and removing nodes of degree 0. Since the scale-free property is consistent with observed sexual contact networks (Schneeberger *et al.* 2004), and for sexually transmitted infections recovered individuals often fail to acquire long-lasting immunity, pairing scale-free networks with SIS dynamics is natural. Epidemic rates were $\tau = 1$, $\gamma = 1$. Figure 2 shows that models incorporating heterogeneity both fit to simulation much better than the homogeneous pairwise model and that the addition of just one extra equation for the heterogeneous pairwise when compared with heterogeneous mixing significantly improves the fit to early growth behaviour.

4. DISCUSSION

We have analysed the conditions under which PGF and heterogeneous-mixing models can be derived from a general pairwise approach. These split into assumptions about the network itself, such as zero clustering (2.3) and non-assortativity (2.5), and assumptions about the interaction of dynamics with network structures such as (2.2), (2.4), (2.6) and (2.7). We have used these conditions to derive a clustered PGF model and a low-dimensional heterogeneous SIS model, which are likely to be of significant utility in the study of epidemics on networks owing to their relatively low dimensionality.

In general, the starting point for analysis of disease transmission on a network has to be the available data. Where the full network is known, or can be imputed with confidence, then explicit stochastic simulation of the epidemic process is likely to be the best approach. When only statistical properties of the network, such as degree distribution d_k, correlation between degrees $\mathcal{C}_{kl}$ or clustering coefficient ϕ, are known, then there are two complementary approaches: either generate exemplar networks for simulation with the appropriate statistics or make use of ODE-based models of the kind considered here, which can be parametrized directly from the network statistics, are numerically tractable and mathematically transparent.

Within this second approach, each ODE-based model used to study network epidemics will have its own domain of validity. While the most general pairwise model can be applied to all compartmental paradigms, with clustering, degree heterogeneity and assortativity included, this comes with increased and potentially unnecessary computational overhead. The more tractable models, on the other hand, may suffer from important inaccuracies when the implicit assumptions they make are not justified.

A question is also posed, however, about the ultimate justification of even the general pairwise model. At present, it is widely believed that in an appropriate regime, this model will be 'exact' in the same way that the standard SIR equations for an epidemic tend to the mean
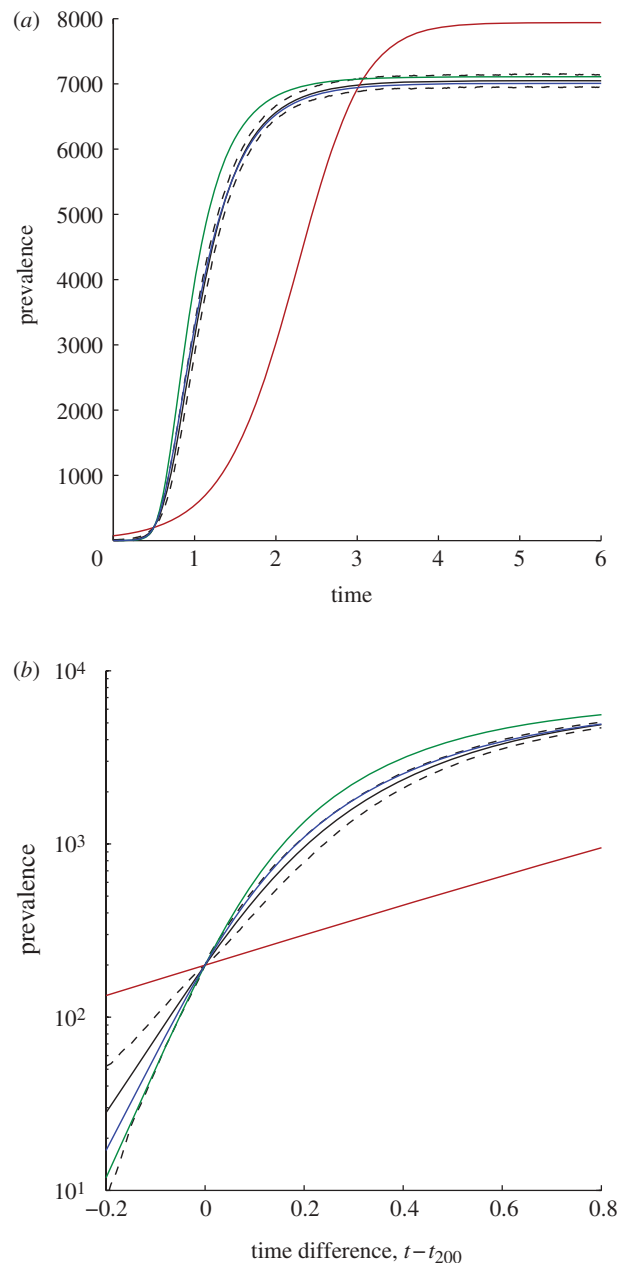

Figure 2. Numerical test of the heterogeneous SIS model and other ODE approaches. The network has size $N \approx 10^4$, and is scale-free with parameters $m_0 = 20$ and $m = 2$. The transmission rate is $\tau = 1.0$ at unit recovery rate. We shift time for each of 10^3 stochastic simulations, so all curves agree on when a prevalence of 200 is reached, and the simulation mean and prediction interval can be meaningfully visualized. Clearly, the heterogeneous pairwise approach is in excellent agreement with simulation. Solid line, simulation mean; dashed line, simulation 95% PI; red line, homogeneous pairwise; green line, heterogeneous mixing; blue line, heterogeneous pairwise.

behaviour of an exact stochastic epidemic model in the appropriately constructed large-population limit. Whether such a regime can be defined remains an important open problem, and one that will hopefully be resolved in the near future.

In conclusion, we hope that this study has clarified the relationship between diverse ODE-based models and extended the repertoire of models available for use.

REFERENCES

Anderson, R. M. & May, R. M. 1992 *Infectious diseases of humans.* Oxford, UK: Oxford University Press.

Bansal, S., Grenfell, B. T. & Meyers, L. A. 2007 When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891. (doi:10.1098/rsif.2007.1100)

Bansal, S., Khandelwal, S. & Meyers, L. A. 2009 Exploring biological network structure with clustered random networks. *BMC Bioinformatics* **10**, 405. (doi:10.1186/1471-2105-10-405)

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

Dangerfield, C. E., Ross, J. V. & Keeling, M. J. 2008 Integrating stochasticity and network structure into an epidemic model. *J. R. Soc. Interface* **6**, 761–774. (doi:10.1098/rsif.2008.0410)

Dieckmann, O. & Heesterbeek, J. A. P. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation.* New York, NY: Wiley.

Eames, K. T. D. 2006 Contact tracing strategies in heterogeneous populations. *Epidemiol. Infect.* **135**, 443–454. (doi:10.1017/S0950268806006923)

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.

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

Ghani, A. C. & Garnett, G. P. 1998 Measuring sexual partner networks for transmission of sexually transmitted diseases. *J. R. Stat. Soc. A* **161**, 227–238. (doi:10.1111/1467-985X.00101)

Ghani, A. C. *et al.* 2009 The early transmission dynamics of H1N1pdm influenza in the United Kingdom. *PLoS Curr. Influenza* **16**, RRN1130.

House, T. & Keeling, M. J. 2010 The impact of contact tracing in clustered populations. *PLoS Comput. Biol.* **6**, e1000721. (doi:10.1371/journal.pcbi.1000721)

Jolly, A. M. & Wylie, J. L. 2002 Gonorrhoea and *chlamydia* core groups and sexual networks in Manitoba. *Sex. Transm. Infect.* **78**(Suppl. 1), i145–i151. (doi:10.1136/sti.78.suppl_1.i145)

Kamp, C. 2009 Untangling the interplay between epidemic spreading and transmission network dynamic. (http://arXiv.org/abs/0912.4189)

Kao, R. R., Danon, L., Green, D. M. & Kiss, I. Z. 2006 Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc. R. Soc. B* **273**, 1999–2007. (doi:10.1098/rspb.2006.3505)

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

Kiss, I. Z., Green, D. M. & Kao, R. R. 2006 The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Math. Biosci.* **203**, 124–136. (doi:10.1016/j.mbs.2006.03.002)

Klovdahl, A. S. 1985 Social networks and the spread of infectious diseases: the AIDS example. *Soc. Sci. Med.* (*1982*), **21**, 1203–1216.

Lysons, R. E., Gibbens, J. C. & Smith, L. H. 2007 Progress with enhancing veterinary surveillance in the United Kingdom. *Vet. Rec.* **160**, 105–112.

Moreno, Y., Pastor-Satorras, R. & Vespignani, A. 2002 Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529.

Mossong, J. *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, 381–391.

Newman, M. E. J. 2003 Mixing patterns in networks. *Phys. Rev. E* **67**, 026126. (doi:10.1103/PhysRevE.67.026126)

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.

Schneeberger, A., Nat, R., Mercer, C. H., Gregson, S. A. J., Ferguson, N. M., Nyamukapa, C. A., Anderson, R. M., Johnson, A. M. & Garnett, G. P. 2004 Scale-free networks and sexually transmitted diseases. *Sex. Transm. Dis.* **31**, 380–387.

Serrano, M. Á. & Boguñá, M. 2006 Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.* **97**, 088701. (doi:10.1103/PhysRevLett.97.088701)

Vernon, M. C. & Keeling, M. J. 2009 Representing the UK's cattle herd as static and dynamic networks. *Proc. R. Soc. B* **276**, 469–476. (doi:10.1098/rspb.2008.1009)

Volz, E. 2008 SIR dynamics in random networks with heterogeneous connectivity. *J. Math. Biol.* **56**, 293–310. (doi:10.1007/s00285-007-0116-4)

Volz, E. & Meyers, L. A. 2007 Susceptible–infected–recovered epidemics in dynamic contact networks. *Proc. R. Soc. B* **274**, 2925–2933. (doi:10.1098/rspb.2007.1159)