Bulletin of Mathematical Biology (2009) 71: 1693–1706
DOI 10.1007/s11538-009-9420-z

ORIGINAL ARTICLE

A Motif-Based Approach to Network Epidemics

Thomas House*, Geoffrey Davies, Leon Danon, Matt J. Keeling

Warwick Mathematics Institute and Dept. Biological Sciences, University of Warwick, Coventry, UK

Received: 18 July 2008 / Accepted: 2 April 2009 / Published online: 25 April 2009

Abstract Networks have become an indispensable tool in modelling infectious diseases, with the structure of epidemiologically relevant contacts known to affect both the dynamics of the infection process and the efficacy of intervention strategies. One of the key reasons for this is the presence of clustering in contact networks, which is typically analysed in terms of prevalence of triangles in the network. We present a more general approach, based on the prevalence of different four-motifs, in the context of ODE approximations to network dynamics. This is shown to outperform existing models for a range of small world networks.

Keywords Infections disease · Modelling · Epidemiology · Pair approximation · Triple approximation · Networks · Motifs

1. Introduction

Mathematical modelling has become an indispensable tool in modern infectious disease epidemiology, with even relatively simple models playing an important role in understanding and controlling outbreaks (Anderson and May, 1992). Increasing resources for numerical computation also mean that it is now possible to incorporate large-scale aggregate data into simulations of epidemics in human and animal populations (Ferguson et al., 2006; Tildesley et al., 2006).

At the same time, and at a lower spatial scale, the structure of contacts between individuals can have an important effect on disease dynamics (Keeling and Eames, 2005), with work in this field most advanced in the modelling of sexually transmitted infections (Eames, 2004; Eames and Keeling, 2002; Service and Blower, 1995). For contacts relevant to the spread of airborne disease, the collection of more comprehensive data is an ongoing project, particularly motivated by concerns over pandemic influenza (Edmunds et al., 1997, 2006; Mossong et al., 2008).

*Corresponding author.
E-mail address: T.A.House@warwick.ac.uk (Thomas House).

Although the lack of fully comprehensive contact network data remains a problem, many observed partial contact networks appear to share common statistical features. In particular, the combination of high clustering amongst nodes and short average path length, commonly known as the small world phenomenon (Watts and Strogatz, 1998), has been observed not only in social networks (Wasserman and Faust, 1994), but also in technological, metabolic and citation networks (Newman, 2003; Strogatz, 2001; Watts, 2003). The use of simple models of network creation that reproduce these statistical features permits us to examine their effects in some detail. The effect of unexpectedly high clustering in networks on disease dynamics has been studied in an ODE-based setting (Keeling, 1999). However, small sub-graphs of higher order, known as *motifs* have also been observed to have significantly different prevalences from those expected in a random case (Milo et al., 2002; Milo, 2004). Until now, the epidemiological effects of varying prevalences of four-motifs or higher have not been considered.

In this paper, we present an ODE-based method for deriving approximate dynamics for epidemics on a network, based on closure of network *SIS* equations at the triple level. We then consider the application of this approach to a small world network, and compare with stochastic simulation.

2. Closure methods for network epidemics

We now present a series of sequentially more complex ODE-based approximations to network epidemic dynamics. The level of detail retained in each approximation corresponds to the sub-network size at which the dynamical equations for the system are closed by approximating larger sub-networks in terms of smaller ones. The first two orders of this process are familiar in the literature, however, they are presented here for methodological and notational clarity.

Although the equations presented here are in terms of *SIS* dynamics for simplicity, our approach is straightforwardly extended to *SIR* or *SEIR* dynamics. Our epidemiological parameters for this system are, throughout, I_0 for the initial number of infectious individuals, τ for the rate at which infection spreads between infectious and susceptible individuals that are linked on a network, and g for the rate at which infectious individuals become susceptible again.

2.1. First-order

Suppose we have a network made of nodes numbered $i = 1, \ldots, N$, with adjacency matrix G defined by

$$G_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are linked,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The mean links per node is $n = \sum_{i,j} G_{ij}/N$. We can then write $[A]$ for the number of network nodes in disease state A and $[A{-}B]$ for the number of pairs of nodes with disease states A and B. We then define a notation for the total numbers of nodes and pairs as below.

$$\big[\bullet\big] := \sum_{A} [A] = N, \qquad \big[-\big] := \sum_{A,B} [A{-}B] = nN. \tag{2}$$

The dynamics of an *SIS*-structured disease process on this network will be given by the following pair of equations, which are exact but unclosed:

$$\frac{d}{dt}[S] = -\tau[S\text{–}I] + g[I], \qquad \frac{d}{dt}[I] = \tau[S\text{–}I] - g[I]. \tag{3}$$

To close this system at first-order, we ignore the pair-level correlations between nodes in given disease states and assume that the number of pairs of a certain type is given by the product of the numbers of nodes of the two types that make up:

$$[A\text{–}B] \approx \left[-\right]\frac{1}{\left[\bullet\right]^2}[A][B]. \tag{4}$$

Substituting the approximation (4) into the exact system (3) yields the familiar 'mean-field' *SIS* equations

$$\frac{d}{dt}[S] = -\frac{\beta}{N}[S][I] + g[I], \qquad \frac{d}{dt}[I] = \frac{\beta}{N}[S][I] - g[I], \tag{5}$$

where $\beta := n\tau$. At this level, we have discarded all information about the underlying network apart from the mean number of connections per node and number of nodes, and have essentially assumed that the population is well mixed.

2.2. *Second order*

We can make a better approximation to the actual behaviour of an epidemic on a network by considering the time evolution of pairs. This will require notation for triples, which can be either closed (triangles) or unclosed (lines). The prevalences of these are written as below

$$\left[\wedge\right] := \sum_{A,B,C}[A\text{–}B\text{–}C], \qquad \left[\triangle\right] := \sum_{A,B,C}[\underline{A\text{–}B\text{–}C}]. \tag{6}$$

An exact set of equations for pairs can then be written down in an analogous way to (3)

$$\begin{aligned}
\frac{d}{dt}[S\text{–}S] &= -2\tau\,[S\text{–}S\text{–}I] + 2g[S\text{–}I],\\
\frac{d}{dt}[S\text{–}I] &= \tau\big([S\text{–}S\text{–}I] - [I\text{–}S\text{–}I] - [S\text{–}I]\big) + g\big([I\text{–}I] - [S\text{–}I]\big),\\
\frac{d}{dt}[I\text{–}I] &= 2\tau\big([I\text{–}S\text{–}I] + [S\text{–}I]\big) - 2g[I\text{–}I],
\end{aligned} \tag{7}$$

where here and throughout this paper we use dotted lines to imply expansion as below:

$$[A\text{–}B\text{–}C] = [A\text{–}B\text{–}C] + [\underline{A\text{–}B\text{–}C}]. \tag{8}$$

We can then close the system by assuming that triple-level prevalences are given by appropriate functions of pair- and node-level variables. The approximations to be substituted

into (8) are, following the argumentation of Keeling (1999),

$$\begin{aligned}[A{-}B{-}C] &\approx \left[\wedge\right]\frac{\left[\bullet\right]}{\left[-\right]^2}\frac{[A{-}B][B{-}C]}{[B]},\\ [\underline{A{-}B{-}C}] &\approx \left[\triangle\right]\frac{\left[\bullet\right]^3}{\left[-\right]^3}\frac{[A{-}B][B{-}C][C{-}A]}{[A][B][C]}.\end{aligned} \tag{9}$$

Our notation is related to the more standard one in which triangle-level clustering is parameterised in terms of a clustering coefficient ϕ by the relations

$$\left[\wedge\right] = (1-\phi)Nn(n-1), \qquad \left[\triangle\right] = \phi Nn(n-1). \tag{10}$$

2.3. Third order

Time evolution of triples through simple closure approximations has been discussed in, for example, Bauch (2005), with the conclusion that it can be preferable to develop more sophisticated pair-level models compared to simple triple-level approaches. Here we consider an approach to triple dynamics that makes use of the full range of possible fourth-order network structures and so avoids the shortcomings of simpler models.

While there are only two connected graphs of degree three, there are six connected graphs of degree four, which we write as

$$\begin{aligned}\left[\nwarrow\right] &:= \sum_{A,B,C,D} [A{-}\overline{B{-}C\ \ D}],\\ \left[\sqcap\right] &:= \sum_{A,B,C,D} [A{-}B{-}C{-}D],\\ \left[\swarrow\right] &:= \sum_{A,B,C,D} [\underline{A{-}B{-}C}{-}D],\\ \left[\square\right] &:= \sum_{A,B,C,D} [\underline{A{-}B{-}C{-}D}],\\ \left[\boxslash\right] &:= \sum_{A,B,C,D} [\underline{A{-}\overline{B{-}C}{-}D}],\\ \left[\boxtimes\right] &:= \sum_{A,B,C,D} [\underline{\underline{A{-}B{-}C}{-}D}].\end{aligned} \tag{11}$$

Exact equations for the time evolution of the triple variables can then be written down in the same way as (3) and (8). We have to consider closed and unclosed triples separately, due to the different possibilities for surrounding motifs, in addition to the internal dynamics of a closed triple, which allow transmission of infection between any of its component nodes. The six equations for unclosed triples and four equations for closed triples are presented below, making use of the notation developed above.

Unclosed triple equations

$$
\begin{aligned}
\frac{d}{dt}[S\text{–}S\text{–}S] &= -\tau\big(2[S\text{–}\overset{\cdot\cdot\cdot\cdot}{S}\text{–}S\text{–}I] + [S\text{–}\overline{S\text{–}S\cdots I}]\big) + g\big(2[S\text{–}S\text{–}I] + [S\text{–}I\text{–}S]\big),\\
\frac{d}{dt}[S\text{–}S\text{–}I] &= \tau\big([S\text{–}\overset{\cdot\cdot\cdot\cdot}{S}\text{–}S\text{–}I] - [S\text{–}\overline{S\text{–}I\cdots I}] - [S\text{–}\overline{S\text{–}I\cdots I}] - [S\text{–}S\text{–}I]\big)\\
&\quad + g\big([S\text{–}I\text{–}I] + [I\text{–}S\text{–}I] - [S\text{–}S\text{–}I]\big),\\
\frac{d}{dt}[S\text{–}I\text{–}S] &= +\tau\big([S\text{–}\overline{S\text{–}S\cdots I}] - 2[S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\text{–}S\text{–}I] - 2[S\text{–}I\text{–}S]\big)\\
&\quad + g\big(2[S\text{–}I\text{–}I] - [S\text{–}I\text{–}S]\big),\\
\frac{d}{dt}[S\text{–}I\text{–}I] &= \tau\big([S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\text{–}S\text{–}I] + [S\text{–}\overline{S\text{–}I\cdots I}] - [S\text{–}\overline{I\text{–}I\cdots I}] + [S\text{–}I\text{–}S]\\
&\quad + [S\text{–}S\text{–}I] - [S\text{–}I\text{–}I]\big) + g\big([I\text{–}I\text{–}I] - 2[S\text{–}I\text{–}I]\big),\\
\frac{d}{dt}[I\text{–}S\text{–}I] &= \tau\big(2[S\text{–}\overline{S\text{–}I\cdots I}] - [I\text{–}\overline{S\text{–}I\cdots I}] - 2[I\text{–}S\text{–}I]\big)\\
&\quad + g\big([I\text{–}I\text{–}I] - 2[I\text{–}S\text{–}I]\big),\\
\frac{d}{dt}[I\text{–}I\text{–}I] &= \tau\big(2[S\text{–}\overline{I\text{–}I\cdots I}] + [I\text{–}\overline{S\text{–}I\cdots I}] + 2[S\text{–}I\text{–}I] + 2[I\text{–}S\text{–}I]\big)\\
&\quad - 3g[I\text{–}I\text{–}I].
\end{aligned}
\tag{12}
$$

Closed triple equations

$$
\begin{aligned}
\frac{d}{dt}[\underline{S\text{–}S\text{–}S}] &= -3\tau[\underline{S\text{–}\overset{\cdot\cdot\cdot\cdot}{S}\text{–}S\text{–}I}] + 3g[\underline{S\text{–}S\text{–}I}],\\
\frac{d}{dt}[\underline{S\text{–}S\text{–}I}] &= \tau\big([\underline{S\text{–}\overset{\cdot\cdot\cdot\cdot}{S}\text{–}S\text{–}I}] - 2[\underline{S\text{–}S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\cdots I}] - 2[\underline{S\text{–}S\text{–}I}]\big)\\
&\quad + g\big(2[\underline{S\text{–}I\text{–}I}] - [\underline{S\text{–}S\text{–}I}]\big),\\
\frac{d}{dt}[\underline{S\text{–}I\text{–}I}] &= \tau\big(2[\underline{S\text{–}S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\cdots I}] - [\underline{S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\text{–}I\cdots I}] + 2[\underline{S\text{–}S\text{–}I}] - 2[\underline{S\text{–}I\text{–}I}]\big)\\
&\quad + g\big([\underline{I\text{–}I\text{–}I}] - 2[\underline{S\text{–}I\text{–}I}]\big),\\
\frac{d}{dt}[\underline{I\text{–}I\text{–}I}] &= 3\tau\big([\underline{S\text{–}\overset{\cdot\cdot\cdot\cdot}{I}\text{–}I\cdots I}] + 2[\underline{S\text{–}I\text{–}I}]\big) - 3g[\underline{I\text{–}I\text{–}I}].
\end{aligned}
\tag{13}
$$

Dotted lines are throughout expanded similarly to (8), as a sum over no line and a full line, so that, for example, the first such term above expands to

$$
[S\text{–}\overset{\cdot\cdot\cdot\cdot}{S}\text{–}S\text{–}I] = [S\text{–}S\text{–}S\text{–}I] + [\underline{S\text{–}S\text{–}S\text{–}I}] + [S\text{–}\overline{S\text{–}S\text{–}I}] + [\underline{S\text{–}\overline{S\text{–}S\text{–}I}}]. \tag{14}
$$

In order to close the system at triple level, we approximate the four-motifs in terms of triples, pairs and nodes in a way that incorporates as much of the internal structure of each four-motif as possible. The approximation scheme that we substitute into equations

(13) and (14) to provide a closed system of triplewise equations is thus:

$$
\begin{aligned}
[A{-}\overline{B{-}C\ \ D}] &\approx \left[\nwarrow\right]\frac{\left[-\right]^3}{\left[\wedge\right]^3\left[\bullet\right]}\frac{[A{-}B{-}C][A{-}B{-}D][C{-}B{-}D][B]}{[A{-}B][B{-}C][B{-}D]},\\
[A{-}B{-}C{-}D] &\approx \left[\sqcap\right]\frac{\left[-\right]}{\left[\wedge\right]^2}\frac{[B{-}C{-}D][A{-}B{-}C]}{[B{-}C]},\\
[\underline{A{-}B{-}C}{-}D] &\approx \left[\swarrow\right]\frac{\left[-\right]^3}{\left[\triangle\right]\left[\wedge\right]^2\left[\bullet\right]}\frac{[\underline{A{-}B{-}C}][B{-}C{-}D][A{-}C{-}D][C]}{[A{-}C][B{-}C][C{-}D]},\\
[\underline{A{-}B{-}C{-}D}] &\approx \left[\square\right]\frac{\left[-\right]^4}{\left[\wedge\right]^4}\frac{[A{-}B{-}C][B{-}C{-}D][C{-}D{-}A][D{-}A{-}B]}{[B{-}C][C{-}D][D{-}A][A{-}B]},\\
[\underline{A{-}\overline{B{-}C}{-}D}] &\approx \left[\boxslash\right]\frac{\left[-\right]}{\left[\triangle\right]^2}\frac{[\underline{B{-}C{-}D}][\underline{A{-}B{-}C}]}{[B{-}C]},\\
[\underline{\underline{A{-}\overline{B{-}C}{-}D}}] &\approx \left[\boxtimes\right]\frac{\left[-\right]^6}{\left[\triangle\right]^4\left[\bullet\right]^4}\\
&\quad\times\frac{[\underline{\underline{A{-}B{-}C}}][\underline{\underline{B{-}C{-}D}}][\underline{\underline{C{-}D{-}A}}][\underline{\underline{D{-}A{-}B}}][A][B][C][D]}{[B{-}C][C{-}D][D{-}A][A{-}B][A{-}C][B{-}D]}.
\end{aligned}
\tag{15}
$$

Each of the approximate, closed systems derived in this section can, given appropriate initial conditions, be numerically integrated using standard techniques for ODEs. We expect them to provide a sequentially better approximation to the expected dynamics of *SIS* epidemics on large contact networks, a hypothesis that can be tested through comparison with stochastic simulation.

3. Application to small world networks

The major benefit of considering higher-order closure of the type we have derived is a more sophisticated treatment of clustering. We therefore consider a Watts–Strogatz small world network (Watts and Strogatz, 1998), which interpolates between a lattice and an Erdös–Rényi random graph (Erdös and Rényi, 1961). The interpolation parameter p represents the probability that any given lattice link is broken and replaced with a random link. Two examples of such networks are shown in Fig. 1. All underlying lattices considered here are toroidal (meaning boundary conditions are periodic).

3.1. Motif prevalences

In order to provide a set of network parameters to integrate the equation sets (3), (8), (13) and (14), we need a set of prevalences for motifs at each level required by the relevant

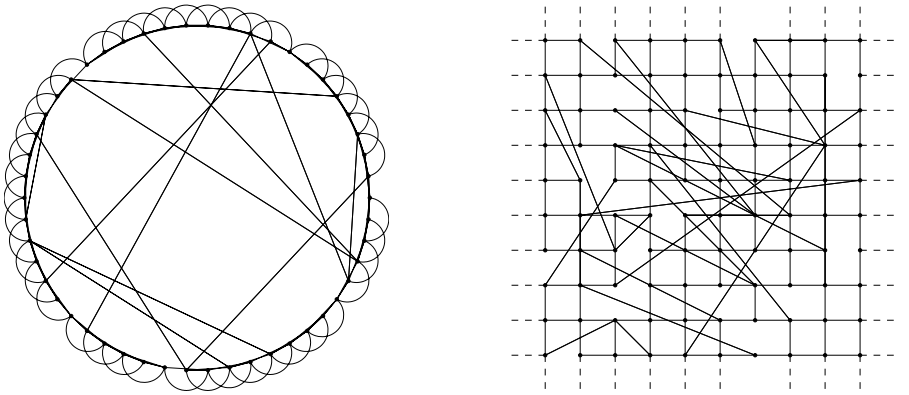

Fig. 1 Two small world networks. On the left, a network based on a one-dimensional lattice of degree two, and on the right, a network based on a two-dimensional lattice of degree one. Both networks have randomisation parameter $p = 0.1$ and mean links per node $n = 4$, however, the variation in underlying lattice structure (reflected in prevalences of higher order motifs) causes these networks to behave very differently in an epidemiological setting.

equation set. These are required at one level above the dynamical variables, to ensure that the closure approximations are correctly normalised.

For the lattices forming the base of the small world networks that we are considering, these prevalences can be calculated using combinatorial arguments. For example, in a 2-d lattice, each node can be associated with eight squares through counting four initial directions clockwise and anticlockwise. Using this and similar arguments, we derived the results below for a 1-d lattice with k-nearest neighbour links (which we call *degree* k) and for a 2-d square lattice with nearest-neighbour links (degree 1).

Motif	1-d lattice	2-d lattice
$[\bullet]$	N	N
$[-]$	$2kN$	$4N$
$[\wedge]$	$k(k+1)N$	$12N$
$[\triangle]$	$3k(k-1)N$	0
$[\llcorner]$	0	$24N$
$[\sqcap]$	$\frac{1}{3}k(k+1)(2k+1)N$	$28N$
$[\boxminus]$	$\frac{2}{3}k(k+1)(k-1)N$	0
$[\square]$	0	$8N$
$[\boxplus]$	$\frac{2}{3}k(k+1)(k-1)N$	0
$[\boxtimes]$	$4k(k-1)(k-2)N$	0

As we introduce a probability p of breaking lattice links and making random ones to generate a small world network, we modify the motif prevalences above. In general, the number of new triangles formed through such a process is proportional to $1/N$ and so in the infinite population size limit (which is implicit in ODE approaches) triangles are only destroyed by randomisation. The same will be true for all other small closed loops, and together with identities relating different motifs, this makes calculation of the effects of randomisation possible.

The arguments above lead to the general set of relations below for a set of motif prevalences following randomisation (subscript p) based on unrandomised prevalences (subscript 0).

$$
\begin{aligned}
\left[\bullet\right]_p &= \left[\bullet\right]_0,\\
\left[-\right]_p &= \left[-\right]_0,\\
\left[\wedge\right]_p &= \left[\wedge\right]_0 + \left(1-(1-p)^3\right)\left[\triangle\right]_0,\\
\left[\triangle\right]_p &= (1-p)^3\left[\triangle\right]_0,\\
\left[\boxtimes\right]_p &= (1-p)^6\left[\boxtimes\right]_0,\\
\left[\boxbslash\right]_p &= (1-p)^5\left[\boxbslash\right]_0 + p(1-p)^5\left[\boxtimes\right]_0,\\
\left[\square\right]_p &= (1-p)^4\left[\square\right]_0 + p(1-p)^4\left[\boxbslash\right]_0 + p^2(1-p)^4\left[\boxtimes\right]_0,\\
\left[\diagdown\right]_p &= (1-p)^3\left[\diagdown\right]_0 + 2\left(2p(1-p)^4 + p^2(1-p)^3\right)\left[\boxbslash\right]_0\\
&\quad + \left(p(1-p)^5 + 3p^2(1-p)^4 + p^3(1-p)^3\right)\left[\boxtimes\right]_0,\\
\left[\sqcap\right]_p &= (n-1)\left[\wedge\right]_p - \left[\diagdown\right]_p - \left[\square\right]_p - \left[\boxbslash\right]_p,\\
\left[\nwarrow\right]_p &= Nn(n-1)(n-2) - 3\left[\diagdown\right]_p - 3\left[\boxbslash\right]_p - \left[\boxtimes\right]_p.
\end{aligned}
\tag{16}
$$

3.2. *Comparison with simulation*

3.2.1. *Simulation details*

The series of sequentially improved ODE models as described above are designed to provide an approximation to a Markovian process of infection and recovery on a network as defined in Appendix A.1.. The numerical integration of such a process involves $2^N - 1$ independent dynamical variables for *SIS* dynamics on a network of size N, in contrast to the ODE approaches outlined which simply incorporate N as a parameter. Since the integration of the full Markov process as N becomes sufficiently large for ODE-based approaches to be appropriate is impossible, we compare our results with stochastic realisations of an appropriate discrete-time Markov chain as defined in Appendix A.2..

For each parameter set, we generated a network through random rewiring (Watts and Strogatz, 1998) and ran 10^4 epidemics, ignoring realisations that ended with no infectious individuals. We used the full simulated data to plot both the mean number of infectious

individuals and the 95% confidence interval around this mean, and then compared these to our ODE models.

3.2.2. Accuracy of ODE approaches

The comparison between the three levels of ODE approaches above and stochastic simulations are shown in Figs. 2 and 3. In each figure, the number of infected individuals is shown over time for our three ODE approaches and also for stochastic simulations, with the full set of parameters involved indicated in the figure captions.

Figure 2 demonstrates the improvements in capturing the average behaviour of a stochastic system that come from the triplewise approach as compared to both pairwise and mean-field approaches on a one-dimensional small-world network of degree 2. We find that for faster epidemics, this improvement is most marked when considering the early growth and transient behaviour of the system, while for slower epidemics the improvement is greatest when considering the endemic state. This is because during early growth of quickly spreading infection in a clustered system, 'bottlenecking' significantly reduces the incidence rate of infection compared to the mean-field R_0, whereas for slower epidemics it is long-term behaviour that is more strongly influenced by clustering, and these facts are captured at increasing levels of detail by both pair- and triple-based approaches.

The implications in different regions of network parameter space are shown in Fig. 3. The plot for a one-dimensional lattice ($p = 0$) shows, as we would expect, that for this system the ODE-based approaches are not appropriate since the growth in infection is initially linear (although it should be noted that they give good approximations to the endemic state). For the $p = 1$ random graph, both the triplewise and pairwise approach give a good approximation to the system with little to choose between them; in this case both still significantly outperform the mean-field model.

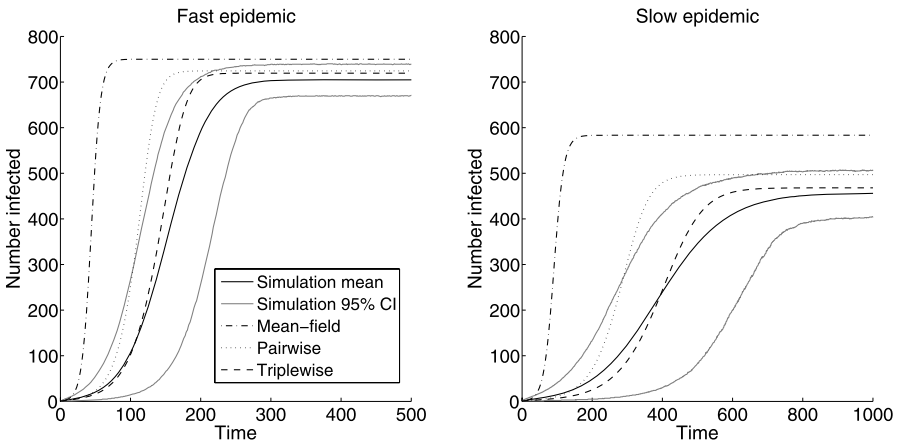

Fig. 2 An *SIS* epidemic with varying transmissibility on a small world network. The total number of nodes is $N = 1000$ and the initial number of infectious individuals is 1. The base network is a one-dimensional lattice of degree 2, with randomisation parameter $p = 0.1$. The disease parameters are $\tau = 0.05$, $g = 0.05$ for the 'Fast epidemic' and $\tau = 0.03$, $g = 0.05$ for the 'Slow epidemic'. For the simulations, we show mean and 95% confidence intervals (defined as the range of numbers infected within which 95% of simulations sit at a given time) for 10^4 runs.

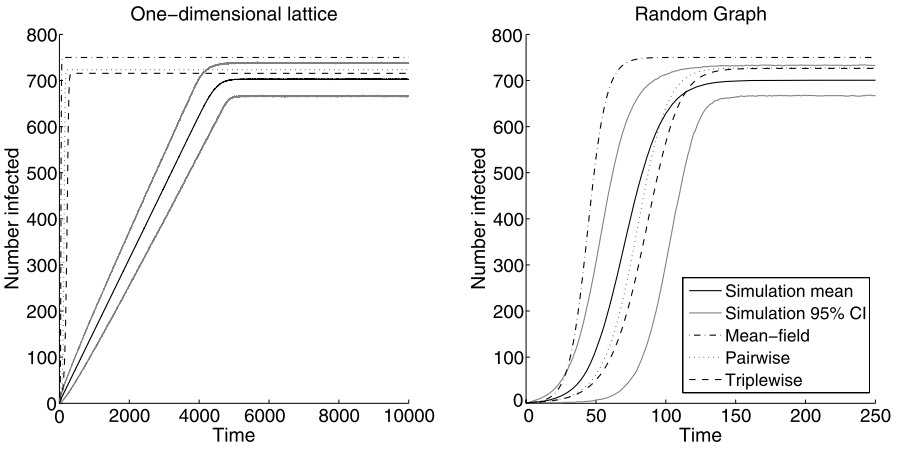

Fig. 3 An *SIS* epidemic on networks with extreme values of the randomisation parameter. The total number of nodes is $N = 1000$ and the initial number of infectious individuals is 1. The base networks are either a one-dimensional lattice of degree 2, (corresponding to randomisation parameter $p = 0$) or else a random graph (corresponding to randomisation parameter $p = 1$). The disease parameters are $\tau = 0.05$, $g = 0.05$. For the simulations, we show mean and 95% confidence intervals (defined as the range of numbers infected within which 95% of simulations sit at a given time) for 10^4 runs.

3.2.3. The two-dimensional lattice

Finally, we investigate a small world network based on a two-dimensional lattice, as shown in Fig. 4. As can be seen in the left-hand plot, in this case the triplewise model provides a significantly improved transient behaviour when compared to the pairwise, spending longer carrying out the non-exponential growth in the numbers infectious seen on the lattice before a period of exponential growth. This is to be expected, since a two-dimensional lattice is clustered at the level of squares but not triangles, a feature incorporated into the triplewise model but not the pairwise.

Once a small world network is created by randomisation, as in the right-hand plot of Fig. 4, the results are not significantly different from the networks based on a one-dimensional lattice. Indeed, in the limit $p \to 1$ we recover the random graph of Fig. 3, with behaviour at intermediate values of p smoothly interpolating between the two-dimensional lattice and random graph. It is worth, at this stage, making some more general points about comparison of stochastic and deterministic models.

For *SIS* dynamics, there are two fixed points of the system: disease-free and endemic. In a deterministic model, the disease-free state is unstable while the disease-free state is stable, when considering small perturbations. For a stochastic model, however, there is always the possibility of stochastic effects leading to the extinction of infection. This creates a dilemma for comparisons between deterministic and stochastic models; if we average over realisations that include stochastic extinction, then we will not obtain a good measure of the quasi-stationary endemic state; but if we ignore such realisations, then our measure of early behaviour will not be correctly averaged.

We therefore consider ODE approximations that lie within the 95% confidence intervals of the stochastic model to be a good fit, and do not attempt to distinguish between them. As an example of why such distinction would be problematic, for the two-

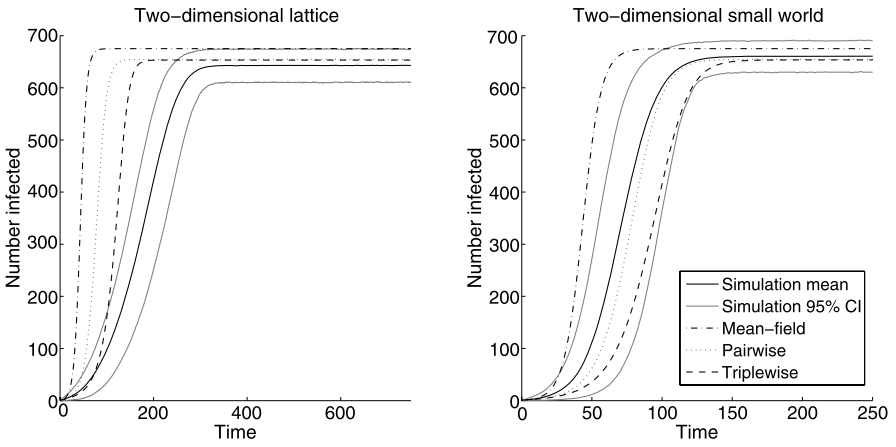

Fig. 4 An *SIS* epidemic on networks based on two-dimensional lattices. The total number of nodes is $N = 900$ and the initial number of infectious individuals is 1. The base networks are either a two-dimensional lattice of degree 1, (corresponding to randomisation parameter $p = 0$) or else a small world network based on a two-dimensional lattice of degree 1, with randomisation parameter $p = 0.1$. The disease parameters are $\tau = 0.05$, $g = 0.05$. For the simulations, we show mean and 95% confidence intervals (defined as the range of numbers infected within which 95% of simulations sit at a given time) for 10^4 runs.

dimensional small world of Fig. 4, while the pairwise prediction is often 'closer' to the simulation mean, the triplewise curve would have better overlap if it were left-shifted, and for the technical reasons noted above it is not clear which of these is 'better'. The mean-field prediction for this plot is clearly significantly different from the overwhelming majority of stochastic realisations, and so both other ODE methods can be meaningfully considered more accurate.

This lack of significant difference leads to an important point about the appropriate way to consider higher order clustering in epidemic models, namely that we need to consider both the full set of four-motif weights, and the extent to which these differ from the weights predicted by models incorporating only lower levels of local structure. So while a 2-d small world network has square-level local structure, this is not manifested in sufficiently distinctive four-motif weights to register a strong epidemiological signature.

4. Discussion

ODE models based on closure of network structure at different orders are a useful tool in understanding the effects of contact networks on disease transmission. We have presented here a triple-level model that incorporates more detailed information about local network structure to improve the agreement between ODE approaches and the output from Monte Carlo realisations of network *SIS* epidemics.

It should be noted that in comparing ODE models and simulation above, we have made no attempt to fit any parameters, but have put in the same values for τ, g, I_0, N, n, p and motif prevalences to each model. This makes the close fit of the triplewise approach in

the appropriate region of parameter space even more remarkable. Indeed, for some of the scenarios considered, the stochastic nature of the comparison simulations is the main limiting factor in determining the accuracy of the ODE models, suggesting that more sophisticated methods for such comparison should be developed.

One particular application for this approach would be to consider realistic networks, where it is likely that triangle-level clustering will be a poor way of characterising the local patterns present in the network; in contrast, the six four-motifs should provide significant extra information about local network properties. This could be particularly important when modelling sexually transmitted infections, where clustering has been shown to impact significantly on the efficacy of contact tracing (Eames and Keeling, 2003), and yet triangles will not be highly prevalent in populations with a dominant heterosexual component.

For this to be successful, it will also be necessary to incorporate heterogeneities in the number of contacts at each node. Similarly, it would be useful to be able to perform the analysis above for an increased number of disease compartments to consider the interaction of generalised notions of clustering with, for example, long-lasting immunity, latent periods, traced and quarantined individuals. While the procedure for deriving such models is a simple generalisation of our method, extra model compartments would render any calculation extremely tedious, suggesting that a priority for future research is to automate the generation of ODEs approximating network epidemics.

Acknowledgements

Work supported by EU grant INFTRANS (FP6 STREP; contract no. 513715). T.H. was supported by a Wellcome Trust VIP Fellowship.

Appendix A: Description of network *SIS* as a Markov process

A.1. Exact formulation

We provide here a full description of *SIS* dynamics on a network of size N as a Markov process in which each possible state of the system can be represented as a member of

$$\chi = \left\{ x \in [0,1]^{2^N} \;\middle|\; \|x\| = 1 \right\}. \tag{A.1}$$

The system is described at any one time by an element $p \in \chi$, which evolves over time according to the master (also called Kolmogorov) equation

$$\frac{dp}{dt} = Qp. \tag{A.2}$$

We define the operator Q in the representation of the system where p is a rank-N tensor, i.e. $p_{A_1 \ldots A_N} = \Pr(\text{node 1 is in state } A_1, \ldots, \text{and node } N \text{ is in state } A_N)$. The full system is described in this representation by the following equations

$$\begin{aligned}\frac{d}{dt}p_{A_1\ldots A_N} &= \sum_{\{B\}} Q_{A_1\ldots A_N}{}^{B_1\ldots B_N} p_{B_1\ldots B_N},\\ Q_{A_1\ldots A_N}{}^{B_1\ldots B_N} &:= \sum_i \left[\tau\delta_S^{B_i}\sum_k \left(G_{ki}\delta_I^{B_k}\right)\left(\delta_{A_i}^I - \delta_{A_i}^S\right) + g\delta_I^{B_i}\left(\delta_{A_i}^S - \delta_{A_i}^I\right)\right]\prod_{j\neq i}\delta_{A_j}^{B_j}.\end{aligned} \tag{A.3}$$

The quantities associated with these equations are defined as follows: δ_A^B takes the value 1 when $A = B$ and the value 0 otherwise. G_{ij} is the network matrix. τ is the rate of infection across a network link, and g is the recovery rate of the disease.

A.2. Discrete time simulation

Given an appropriately small timestep, δt, we can simulate the behaviour of the system of Section A.1. through choosing a random number r_i uniformly distributed over the interval $[0, 1]$ for each node i. The infectious state of each node is then modified under the following conditions:

$$\begin{aligned} I_i(t) &\to S_i(t+\delta t) \quad \text{if } r_i > 1 - g\delta t,\\ S_i(t) &\to I_i(t+\delta t) \quad \text{if } r_i > (1-\tau\delta t)^{\sum_j \delta_{A_j(t),I} G_{ji}},\end{aligned} \tag{A.4}$$

where we use the notation $A_i(t) \in \{S, I\}$ to represent the state of node i at time t, and the Kronecker delta to translate this into a value of either 1 or 0. In our simulations, we set $\delta t = 1$ and keep the values of τ, g small.

Appendix B: Useful identities

For *SIS* dynamics, the unclosed triples are related to each other through

$$\left[\wedge\right] = [S\text{–}S\text{–}S] + 2[S\text{–}S\text{–}I] + [S\text{–}I\text{–}S] + 2[S\text{–}I\text{–}I] + [I\text{–}S\text{–}I] + [I\text{–}I\text{–}I]. \tag{B.1}$$

The equivalent equation for closed triples is

$$\left[\triangle\right] = [\underline{S\text{–}S\text{–}S}] + 3[\underline{S\text{–}S\text{–}I}] + 3[\underline{S\text{–}I\text{–}I}] + [\underline{I\text{–}I\text{–}I}]. \tag{B.2}$$

The following identities should hold on any network composed of a 'giant node' (i.e. with a path through the network linking any two of its nodes).

$$\begin{aligned} \left[\wedge\right] + \left[\triangle\right] &= Nn(n-1),\\ \left[\leftthreetimes\right] + 2\left[\boxslash\right] + \left[\boxtimes\right] &= (n-2)\left[\triangle\right],\\ \left[\nwarrow\right] + 2\left[\leftthreetimes\right] + \left[\boxslash\right] &= (n-2)\left[\wedge\right],\\ \left[\sqcap\right] + \left[\leftthreetimes\right] + \left[\square\right] + \left[\boxslash\right] &= (n-1)\left[\wedge\right].\end{aligned} \tag{B.3}$$

References

Anderson, R.M., May, R.M., 1992. Infectious Diseases of Humans. Oxford University Press, Oxford.
Bauch, C., 2005. The spread of infectious diseases in spatially structured populations: An invasory pair approximation. Math. Biosci. 198(2), 217–237.
Eames, K., 2004. Monogamous networks and the spread of sexually transmitted diseases. Math. Biosci. 189(2), 115–130.
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, 13330–13335.
Eames, K.T.D., Keeling, M.J., 2003. Contact tracing and disease control. Proc. R. Soc. Lond. B, Biol. 270, 2565–2571.
Edmunds, W.J., et al., 1997. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc. R. Soc. Lond. B, Biol. 264(1384), 949–957.
Edmunds, W.J., et al., 2006. Mixing patterns and the spread of close-contact infectious diseases. Emerg. Themes Epidemiol. 3, 10.
Erdös, P., Rényi, A., 1961. On the evolution of random graphs. Bull. Inst. Int. Stat. 38, 343–347.
Ferguson, N.M., et al., 2006. Strategies for mitigating an influenza pandemic. Nature 442(7101), 448–452.
Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. B, Biol. Sci. 266(1421), 859–867.
Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. J. R. Soc. Interface 2(4), 295–307.
Milo, R., 2004. Superfamilies of evolved and designed networks. Science 303(5663), 1538–1542.
Milo, R., et al., 2002. Network motifs: Simple building blocks of complex networks. Science 298(5594), 824–827.
Mossong, J., et al., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. Plos Med. 5(3), 381–391.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45(2), 167–256.
Service, S.K., Blower, S.M., 1995. HIV transmission in sexual networks—an empirical analysis. Proc. R. Soc. Lond. B, Biol. 260(1359), 237–244.
Strogatz, S.H., 2001. Exploring complex networks. Nature 410(6825), 268–276.
Tildesley, M.J., et al., 2006. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 440(7080), 83–86.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.
Watts, D.J., 2003. Six Degrees: The Science of a Connected Age. Norton, New York.
Watts, D., Strogatz, S., 1998. Collective dynamics of 'small-world' networks. Nature 393(6684), 440–442.