Bulletin of Mathematical Biology (2009) 71: 888–905
DOI 10.1007/s11538-008-9386-2

ORIGINAL ARTICLE

A Contact-Network-Based Formulation of a Preferential Mixing Model

Istvan Z. Kiss[a,*], Péter L. Simon[b], Rowland R. Kao[c]

[a] *Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9RF, UK*
[b] *Institute of Mathematics, Eötvös Loránd University Budapest, Budapest, Hungary*
[c] *Faculty of Veterinary Medicine, Institute of Comparative Medicine, University of Glasgow, Glasgow G61 1QH, UK*

Received: 12 June 2008 / Accepted: 2 December 2008 / Published online: 12 February 2009

Abstract Heterogeneity in the number of potentially infectious contacts and connectivity correlations ("like attaches to like", i.e., assortatively mixed or "opposites attract", i.e., disassortatively mixed) have important implications for the value of the basic reproduction ratio R_0 and final epidemic size. In this paper, we present a contact-network-based derivation of a simple differential equation model that accounts for preferential mixing based on the number of contacts. We show that results based on this model are in good qualitative agreement with results obtained from preferential mixing models used in the context of sexually transmitted diseases (STDs). This simple model can accommodate any mixing pattern ranging from completely disassortative to completely assortative and allows the derivation of a series of analytical results.

Keywords Preferential mixing · Epidemics · Final epidemic size

1. Introduction

Models of infectious disease transmission often assume homogeneous random mixing. This implies that all individuals are equally likely to contact each other and, therefore, if infected, are equally likely to infect susceptible members of the population. The availability of more accurate data at the individual level, the collection of which has been partially driven by the epidemics of HIV/AIDS (Liljeros et al., 2001; Jones and Handcock, 2003; Catania et al., 1992; Anderson et al., 1990; Gupta et al., 1989), SARS (Lipsitch et al., 2003; Hufnagel et al., 2004; Meyers et al., 2005), Foot-and-Mouth Disease (FMD; Ferguson et al., 2001; Keeling et al., 2001; Kiss et al., 2005, 2006a; Green et al., 2006; Kao et al., 2006b), and the possibility of a world wide Pandemic Influenza (Eubank et al., 2004; Ferguson et al., 2005), has highlighted the important role played by contact heterogeneity,

*Corresponding author.
E-mail address: i.z.kiss@sussex.ac.uk (Istvan Z. Kiss).

spatial structure, and connectivity correlations. Differential-equation-based models can be adapted to capture such properties and are amenable to a variety of powerful analytical tools (Kiss et al., 2006b). Hethcote and Yorke (1984) proposed a model for the transmission of gonorrhea that distinguishes between very active and active individuals and defines a mixing matrix that interpolates between the case of proportionate mixing and the case when mixing or contact exclusively happens within the different groups. However, their study was not aimed at investigating the implications of different mixing patterns for epidemic dynamics and most of their results are based on the numerical integration of a system of eight differential equations.

In the context of modeling sexually transmitted diseases (STDs) (Anderson and Garnett, 2000), individuals and their interactions are often conveniently modeled as a network where individuals are represented by nodes and potentially infectious contacts by links between the nodes (Kretzschmar et al., 1996; Ghani et al., 1997). This approach allows to accommodate contact heterogeneity and connectivity correlations in a straightforward way. The network of sexual contacts has important implications for the dynamics of the spread and control of STDs. For example, many control policies for STDs rely on the identification and control of 'core groups' of individuals with a large numbers of interactions (Hethcote and Yorke, 1984; Kretzschmar et al., 1996). Connectivity correlations within the contact network are equally important. The assortativeness of mixing ("like attaches to like") where individuals with similar activity levels are more likely to be in contact leads to a faster initial spread and a smaller overall epidemic while disassortatively mixed ("opposites attract") contact networks lead to a slower initial rise in the epidemic, but a much larger epidemic over a long period of time (Gupta et al., 1989; Anderson et al., 1990; Ghani et al., 1997).

Empirical evidence such as the correlation properties of the Internet (Pastor-Satorras et al., 2001) has also led to the recent development of numerous network- and differential-equation-based models (Boguñá et al., 2003a, 2003b; Newman, 2002, 2003; Moreno et al., 2003; Barthélemy et al., 2005). Many of these models focus on the effect of connectivity correlations on the epidemic threshold, initial growth rate, and hierarchical spread. For example, it has been shown that epidemics on networks characterized by high variance in node degree grow rapidly, and in the limiting case of infinite variance, the growth is instantaneous independently of the mixing pattern (Boguñá et al., 2003a). A key ingredient of differential-equation-based models that account for contact between and within groups of individuals with various levels of activity (i.e., number of contacts) is the mixing matrix. This provides information about the amount of contact within and between groups. In many studies (Anderson et al., 1990), the mixing matrix, that is subject to some constraints, is chosen in some convenient way to reflect a desired mixing pattern. In this paper, starting from a contact network representation of the population, we propose a compartmental model that accounts for preferential mixing based on the number of contacts that individuals have. We extend the simple SIR (susceptible, infectious and recovered) model to the case of two groups with different levels of activity and a mixing pattern that can be tuned to vary from completely disassortative to completely assortative. We show that results based on this simple model are in good qualitative agreement with results obtained from models used in the context of STDs and with other individual-based simulation models (Kiss et al., 2008). We also derive a series of analytical results that are difficult to obtain from more detailed compartmental models or complex individual-based simulations.

2. Model

2.1. Population mixing model

Connectivity correlations are based on the number of contacts that individuals have. Hence, in the simplest case, the population is divided into poorly and highly connected individuals. Let N_L and N_H denote the number of poorly and highly connected individuals with k and σk connections per individual respectively, and with $\sigma > 1$ (Kao, 2006a). The total population size is denoted by $N = N_L + N_H$, and $f = \frac{N_H}{N}$ denotes the proportion of highly connected individuals out of the total population. Hence, $\frac{N_L}{N} = 1 - f$. The average number of connections per individual is given by

$$\langle k \rangle = \frac{kN_L + \sigma k N_H}{N_L + N_H}. \tag{1}$$

If contact would occur at random then the probability of connection between two individuals is proportional to the product of their degree (i.e., proportionate mixing). However, we wish to model the situation when individuals in one particular group may or may not preferentially mix with or contact individuals of the same type. To accommodate this preferential mixing, we assume that a proportion $0 \leq a \leq 1$ of all contacts within the population occur between highly connected individuals. Similarly, let $0 \leq b \leq 1$ represent the proportion of contacts between individuals that are less well connected. Therefore, $0 \leq 1 - a - b \leq 1 (\Rightarrow a + b \leq 1)$ represents the proportion of contacts between poorly and highly connected individuals. This leads to the following definitions for the connection correlations within the population:

$$P(L|L) = \frac{b}{1-a}, \qquad P(H|L) = \frac{1-a-b}{1-a}, \tag{2}$$

and

$$P(L|H) = \frac{1-a-b}{1-b}, \qquad P(H|H) = \frac{a}{1-b}. \tag{3}$$

The conditional probabilities defined in Eqs. (2) and (3) are bounded below by zero and from above by one and they satisfy $P(L|L) + P(H|L) = 1$ and $P(L|H) + P(H|H) = 1$. This means that all links starting at a highly connected individual will connect to poorly connected individuals with probability $P(L|H)$ and to highly connected individuals with probability $P(H|H)$. Based on the connectivity correlations, we can define a mixing matrix that allows us to measure the level and type of mixing within the population

$$\begin{array}{c} \\ L \\ H \end{array} \begin{array}{c} \begin{array}{cc} L & H \end{array} \\ \begin{pmatrix} b & \frac{1-a-b}{2} \\ \frac{1-a-b}{2} & a \end{pmatrix} \end{array} = E. \tag{4}$$

The entries of the matrix (e_{ij}) represent the proportion of connections/links within and between the two different subgroups. Let the sum of the rows and columns be denoted by $\alpha_i = \sum_{j \in \{L,H\}} e_{ij}$ and $\beta_j = \sum_{i \in \{L,H\}} e_{ij}$, respectively, and with $i, j \in \{L, H\}$. To quantify

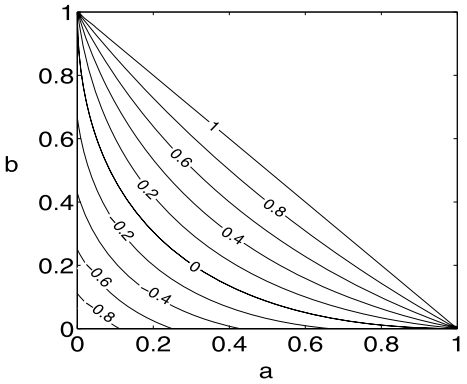

Fig. 1 Illustration of M, defined in Eq. (5), as a function of a and b such that $a + b \leq 1$.

the level of mixing within the population, consider the assortativity coefficient (Newman, 2002, 2003) defined as

$$M = \frac{\sum_{i\in\{L,H\}} e_{ii} - \sum_{i\in\{L,H\}} \alpha_i \beta_i}{1 - \sum_{i\in\{L,H\}} \alpha_i \beta_i} = \frac{a + b - (\frac{1-a+b}{2})^2 - (\frac{1+a-b}{2})^2}{1 - (\frac{1-a+b}{2})^2 - (\frac{1+a-b}{2})^2}$$
$$\in [-1, 1]. \tag{5}$$

The assortativity coefficient expresses the degree to which like connects to like. For random or proportionate mixing, this formula gives $M = 0$, since $e_{ij} = \alpha_i \beta_j$. If contacts happen only between individuals belonging to the same subpopulation (assortatively mixed population), then $\sum_{i\in\{L,H\}} e_{ii} = 1$ and $M = 1$. If contacts happen only between individuals that belong to different subpopulations (disassortatively mixed population), then $e_{ii} = 0$ and $M = -\sum_{i\in\{L,H\}} \alpha_i \beta_i / (1 - \sum_{i\in\{L,H\}} \alpha_i \beta_i)$, which lies in general in the range $-1 \leq M < 0$. In Fig. 1, M is plotted as a function of a and b restricted to $a + b \leq 1$. If $a = b$, then $M = 4a - 1$ and M spans from -1 to 1 along the first diagonal $a = b$. If $a + b = 1$, then $M = 1$ with complete assortativity and with the population fragmented into two noninteracting subpopulations. Different combinations of a and b can result in the same value of M.

2.2. *Disease transmission model*

Individuals from the population are divided into compartments according to one of three states of disease progression: susceptible ($S(\rightarrow S_L, S_H)$); infected and infectious ($I(\rightarrow I_L, I_H)$); and, finally, removed nodes ($R(\rightarrow R_L, R_H)$), which are no longer infectious or they are immune. Within this mixing model, any individual can infect any other individual provided that the mixing model allows for contact. For example, for any choice of a and b, such that $a + b = 1$, it follows that $M = 1$, and in this case the population consists of two noninteracting subpopulations. Therefore, transmission is not possible between poorly and highly connected individuals. Based on these assumptions, we have the following

system of differential equations:

$$\frac{dS_L}{dt} = -\tau k S_L \left(P(L|L)\frac{I_L}{N_L} + P(H|L)\frac{I_H}{N_H} \right), \tag{6}$$

$$\frac{dS_H}{dt} = -\tau\sigma k S_H \left(P(L|H)\frac{I_L}{N_L} + P(H|H)\frac{I_H}{N_H} \right), \tag{7}$$

$$\frac{dI_L}{dt} = \tau k S_L \left(P(L|L)\frac{I_L}{N_L} + P(H|L)\frac{I_H}{N_H} \right) - \gamma I_L, \tag{8}$$

$$\frac{dI_H}{dt} = \tau\sigma k S_H \left(P(L|H)\frac{I_L}{N_L} + P(H|H)\frac{I_H}{N_H} \right) - \gamma I_H, \tag{9}$$

$$\frac{dR_L}{dt} = \gamma I_L, \tag{10}$$

$$\frac{dR_H}{dt} = \gamma I_H, \tag{11}$$

where τ is the per contact rate of transmission and γ is the recovery rate. The right-hand side in Eq. (8) describes the creation of new infections and is proportional to the transmission rate τ, the number of contacts k of poorly connected susceptible individuals, the number of susceptible individuals with k connections, and the probability that any given neighbor of a susceptible individual with k connections is infected. The formulation for such type of heterogeneous contact follows from that of Anderson and May (1991) for sexually transmitted diseases. A similar approach is used when modeling correlated complex networks (Boguñá et al., 2003a, 2003b; Barthélemy et al., 2005). In this present model, following May and Lloyd (2001), it is assumed that contacts switch at a rate that is much faster than the rate at which disease spreads, and thus infection does not result in the loss of one susceptible neighbor upon becoming infectious. Therefore, in this case, the network is dynamic and infection over the network does not result in the loss of a susceptible partner. Equations (6) to (11) are nondimensionalized using N to scale all variables and $1/\gamma$ to scale time. We note that it is sufficient to consider the first four equations since the values of R_L and R_H are determined by the values of the other four variables. Hence, the reduced system is given by

$$\frac{ds_L}{dt} = -\tau s_L(a_{LL}i_L + a_{HL}i_H), \tag{12}$$

$$\frac{ds_H}{dt} = -\tau s_H(a_{LH}i_l + a_{HH}i_H), \tag{13}$$

$$\frac{di_L}{dt} = \tau s_L(a_{LL}i_L + a_{HL}i_H) - i_L, \tag{14}$$

$$\frac{di_H}{dt} = \tau s_H(a_{LH}i_l + a_{HH}i_H) - i_H, \tag{15}$$

where

$$a_{LL} = \frac{kP(L|L)}{\gamma(1-f)}, \qquad a_{HL} = \frac{kP(H|L)}{\gamma f},$$
$$a_{LH} = \frac{k\sigma P(L|H)}{\gamma(1-f)}, \qquad a_{HH} = \frac{k\sigma P(H|H)}{\gamma f}. \tag{16}$$

3. Results

3.1. The basic reproduction number R_0

Following van den Driessche and Watmough (2002), we note that only compartments i_L and i_H are involved in the calculation of R_0. At the disease-free equilibrium $(s_L, s_H, i_L, i_H) = (1-f, f, 0, 0)$ the rate of appearance of new infections F and the rate of transfer of individuals out of the two compartments V are given by

$$F = \begin{array}{c} \\ i_L \\ i_H \end{array} \begin{array}{c} \begin{array}{cc} i_L & i_H \end{array} \\ \begin{pmatrix} \tau k P(L|L) & \frac{\tau k P(H|L)(1-f)}{f} \\ \frac{\tau\sigma k P(L|H) f}{1-f} & \tau\sigma k P(H|H) \end{pmatrix} \end{array}, \qquad V = \begin{array}{c} \\ i_L \\ i_H \end{array} \begin{array}{c} \begin{array}{cc} i_L & i_H \end{array} \\ \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix} \end{array}. \tag{17}$$

The basic reproduction number R_0 is defined as the leading eigenvalue of the next generation matrix FV^{-1}. Solving the resulting quadratic equation the leading eigenvalue, and hence R_0 is given by

$$R_0 = \frac{\tau k}{2\gamma}\Big(P(L|L) + \sigma P(H|H) + \sqrt{\big(P(L|L) - \sigma P(H|H)\big)^2 + 4\sigma P(L|H)P(H|L)}\Big). \tag{18}$$

High values of σ denote high levels of heterogeneity with a considerable difference between the number of connections of individuals in the two subpopulations. Higher number of connections lead to a higher number of individuals being infected by an index case and, therefore, to a higher value of R_0. This is illustrated in Fig. 2 where contour plots of R_0 are given for increasing values of σ while keeping all disease spread and disease progression related parameters constant. We note that varying the proportion of contacts linking highly connected individuals, a, has a higher impact on R_0 compared to varying b. Figure 2 also illustrates how higher levels of assortativity (i.e., in Fig. 2 the region close to $a + b = 1$) lead to higher values of R_0. Therefore, for the same disease, an outbreak is more likely to happen if the population is assortatively mixed. At the individual level highly connected individuals preferentially connect to other highly connected individuals and this leads to a high number of secondary infections generated by an initial index case. The type of the index case (i.e., poorly or highly connected) becomes important in a stochastic formulation where the initial phase in the spread of the disease is crucial. For example, stochastic extinction is more likely to happen if the seeding occurs solely in the poorly connected group. However, in the deterministic continuous formulation, even

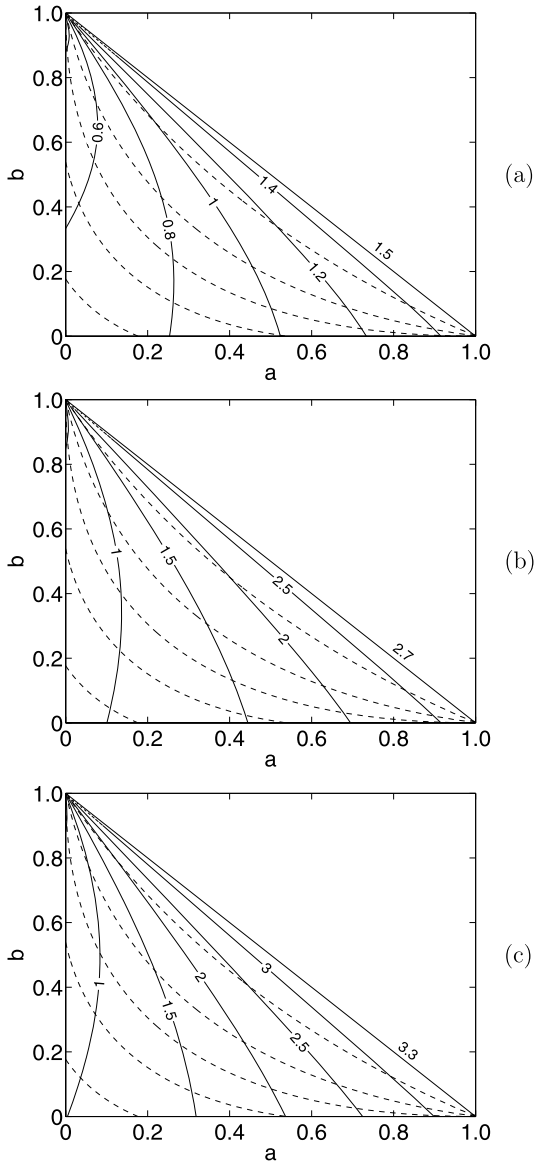

Fig. 2 Based on Eq. (18), R_0 plotted as a function of a and b for $\sigma = 5, 9$, and 11, and $f = 0.25, 0.125$, and 0.1 in (a), (b), and (c), respectively. For all plots, $\tau k/2\gamma = 0.15$. Dotted lines from left to right show the various levels of mixing $M = -0.7, -0.3, 0, 0.3, 0.7$.

weak coupling between the two groups means that an epidemic will be sustained provided $R_0 > 1$.

The transmission potentials within the individual subpopulations can be defined as $\rho_L = \frac{\tau k}{\gamma}$ and $\rho_H = \frac{\tau \sigma k}{\gamma}$. These represent the number of secondary infections produced

by a single infectious individual during its infectious period when introduced into a fully susceptible subpopulation. The basic reproduction number R_0 has a similar definition but the index case is weighted according to the probability of it itself becoming infected. In the case of noninteracting subpopulations the transmission potentials are equivalent to R_0 within the subpopulation. Then R_0 can be written in terms of the transmission potentials to give

$$R_0 = \frac{1}{2}\big(\rho_L P(L|L) + \rho_H P(H|H) + \sqrt{\big(\rho_L P(L|L) - \rho_H P(H|H)\big)^2 + 4\rho_L \rho_H P(L|H)P(H|L)}\big). \tag{19}$$

The limit of $M = -1$ represents a disassortatively mixed population meaning that there are no within subpopulation connections. Therefore, $(P(L|L) = P(H|H) = 0$, $P(L|H) = P(H|L) = 1)$ and R_0 in this case is given by

$$R_0 = \sqrt{\rho_L \rho_H}. \tag{20}$$

This represents the geometric mean of the individual basic reproduction numbers corresponding to individual subpopulations. Hence, a complete cycle of infection from poorly to highly connected individuals and then back again, results in $\rho_L \rho_H$ new infections (Diekmann and Heesterbeek, 2000). If $M = 1$, the population is divided into two distinct and noninteracting subpopulations $(P(L|H) = P(H|L) = 0$, $P(L|L) = P(H|H) = 1)$ and the basic reproduction number is given by

$$R_0 = \frac{1}{2}\max\big(\rho_L + \rho_H \pm (\rho_L - \rho_H)\big). \tag{21}$$

In the current setting, since $\sigma > 1$, the basic reproduction number $R_0 = \rho_H$ since $\rho_H \geq \rho_L$.

3.2. Epidemic dynamics

In the case of $R_0 > 1$, the dynamics of the outbreak is illustrated in Fig. 3 by plotting the proportion of infected (and infectious) individuals (i_L and i_H) as a function of time. These plots are based on numerical solutions of the system of differential Eqs. (12) to (15) using the fourth-order Runge–Kutta method. The relation between R_0 and M (see Fig. 2) allows to choose a and b so that either R_0 or M can be kept constant when comparing different scenarios. The top panel corresponds to the case when R_0 is constant independently of the mixing pattern $M = \pm 0.5$. The lower panel corresponds to the situation where R_0 is not controlled for, but is rather the result of the interaction between disease characteristics and population contact structure. This in effect corresponds to the same disease spreading on populations with different mixing patterns. When the population is assortatively mixed, highly connected individuals become infected at a faster rate than poorly connected individuals. This is also true for the case considered in Fig. 3a even though this case corresponds to a situation where the mixing pattern is adjusted to keep R_0 the same. There is also a significant difference between how quickly an epidemic spreads and how long it lasts for (Fig. 3b). Epidemics on assortatively mixed populations have a fast turnover when compared to the disassortatively mixed case. Epidemics on disassortatively mixed populations are slow to take off and their turnover time is also longer.

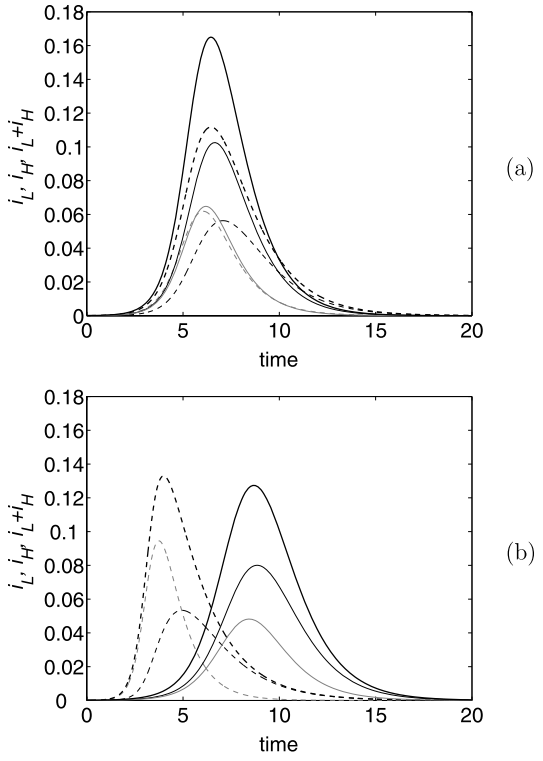

Fig. 3 Proportion of the population infected as a function of time for disassortatively (solid lines) and assortatively (dashed lines) mixed populations. Cumulative proportion (thick lines, $i_L + i_H$) and proportion of poorly (thin black lines, i_L) and highly connected (thin grey lines, i_H) individuals that are infected are all illustrated. For both (a) and (b), $k = 3$, $\sigma = 5$, $f = 0.25$, $\gamma \simeq 0.286$, and $\tau = 0.085$. The level of mixing for disassortatively and assortatively mixed populations is $M = -0.5$ ($a = 0.31, b = 0.01$) and $M = 0.5$ ($a = 0.21, b = 0.57$) (a), and $M = -0.8$ ($a = b = 0.05$) and $M = 0.8$ ($a = b = 0.45$) (b). The basic reproduction number is $R_0 = R_0^{\text{disass}} = R_0^{\text{ass}} = 2.5$ with $r_{\text{disass}}(\infty) \simeq 0.43 + 0.24 = 0.67, r_{\text{ass}}(\infty) \simeq 0.29 + 0.24 = 0.53$ (a), and $R_0^{\text{disass}} = 2.0$ and $R_0^{\text{ass}} = 3.7$ with $r_{\text{disass}}(\infty) \simeq 0.42 + 0.23 = 0.65, r_{\text{ass}}(\infty) \simeq 0.25 + 0.24 = 0.49$ (b).

3.3. Final epidemic size

The severity of an epidemic outbreak is well characterized by the final epidemic size $r(\infty) = r^{\infty} = r_L(\infty) + r_H(\infty) = r_L^{\infty} + r_H^{\infty}$ and we derive analytical results in the form of a family of parametric equations that link the final epidemic size and the two transmission potentials. For ease of notation, Eqs. (12)–(15) are rewritten to give

$$\dot{s}_m = -\tau s_m \sum_{n\in\{L,H\}} a_{nm} i_n, \tag{22}$$

$$\dot{i}_m = \tau s_m \sum_{n\in\{L,H\}} a_{nm} i_n - i_n, \tag{23}$$

with $m \in \{L, H\}$. By using the following notation,

$$\lambda_m = \tau \sum_{n\in\{L,H\}} a_{nm} i_n, \quad m \in \{L, H\} \tag{24}$$

and by integrating Eq. (22) with initial conditions $s_L(0) = N_L/N$ and $s_H(0) = N_H/N$, we obtain

$$s_m(t) = \frac{N_m}{N} \exp\big(-\Phi_m(t)\big), \quad m \in \{L, H\}, \tag{25}$$

where

$$\Phi_m(t) = \int_0^t \lambda_m(s)\, ds, \quad m \in \{L, H\}. \tag{26}$$

Here, for consistency of notation, N_L/N and N_H/N are used instead of $(1 - f)$ and f. By adding Eqs. (22) and (23) and integrating from 0 to ∞, it follows that

$$r_m^\infty = \int_0^\infty i_m(s)\, ds, \quad m \in \{L, H\}. \tag{27}$$

This is obtained upon using $i_m(\infty) = i_m^\infty = 0$ and that $s_m(\infty) = s_m^\infty = \frac{N_m}{N} - r_m^\infty$ for $m \in \{L, H\}$. Considering the limit of $t \to \infty$ in Eq. (26), it follows that

$$\begin{aligned}\Phi_m(\infty) = \Phi_m^\infty &= \tau \int_0^\infty \sum_{n\in\{L,H\}} a_{nm} i_n(s)\, ds \\ &= \tau \sum_{n\in\{L,H\}} a_{nm} \int_0^\infty i_n(s)\, ds = \tau \sum_{n\in\{L,H\}} a_{nm} r_n^\infty.\end{aligned} \tag{28}$$

Taking into account again that $s_m^\infty = \frac{N_m}{N} - r_m^\infty$ for $m \in \{L, H\}$ and using Eq. (25) in the limit of $t \to \infty$, we obtain

$$r_m^\infty = \frac{N_m}{N}\big(1 - \exp\big(-\Phi_m^\infty\big)\big), \quad m \in \{L, H\}, \tag{29}$$

and

$$r^\infty = r_L^\infty + r_H^\infty = \sum_{m\in\{L,H\}} \frac{N_m}{N}\big(1 - \exp\big(-\Phi_m^\infty\big)\big). \tag{30}$$

Combining Eqs. (28) and (29), we obtain an implicit formula for Φ_m^∞ ($m \in \{L, H\}$)

$$\Phi_m^\infty = \tau \sum_{n\in\{L,H\}} a_{nm} \frac{N_n}{N}\big(1 - \exp\big(-\Phi_n^\infty\big)\big), \quad m \in \{L, H\}. \tag{31}$$

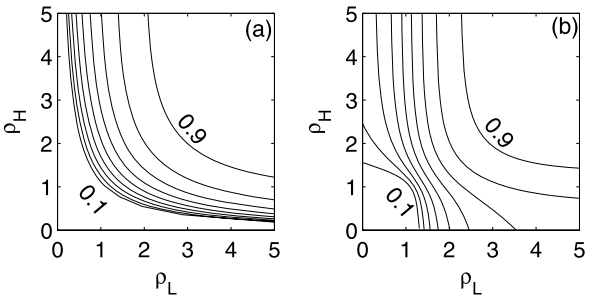

Fig. 4 Final epidemic size as a function of the transmission potentials for (a) disassortatively-mixed ($a = b = 0.05 \rightarrow M = -0.8$) and (b) assortatively-mixed ($a = b = 0.45 \rightarrow M = 0.8$) populations. The case of $f = 0.25$ is considered here.

Combining Eq. (30) and the rearranged version of Eq. (31) and upon using the notations given in Eq. (16), the following parametric equations are obtained:

$$r^{\infty}\left(\Phi_L^{\infty}, \Phi_H^{\infty}\right) = \frac{N_L}{N}\left(1 - \exp\left(-\Phi_L^{\infty}\right)\right) + \frac{N_H}{N}\left(1 - \exp\left(-\Phi_H^{\infty}\right)\right), \tag{32}$$

$$\rho_L\left(\Phi_L^{\infty}, \Phi_H^{\infty}\right) = \frac{\Phi_L^{\infty}}{P(L|L)(1 - \exp(-\Phi_L^{\infty})) + P(H|L)(1 - \exp(-\Phi_H^{\infty}))}, \tag{33}$$

$$\rho_H\left(\Phi_L^{\infty}, \Phi_H^{\infty}\right) = \frac{\Phi_H^{\infty}}{P(L|H)(1 - \exp(-\Phi_L^{\infty})) + P(H|H)(1 - \exp(-\Phi_H^{\infty}))}. \tag{34}$$

In Fig. 4, based on Eqs. (32) to (34), plots of the final epidemic size (r^{∞}) are given for a range of transmission potentials for disassortatively- and assortatively-mixed populations. We separately consider the case of small and large final epidemic size and discuss the dependence of r^{∞} on the values of the transmission potentials. For assortatively-mixed populations (Fig. 4b), an initial perturbation results in small final epidemic sizes for transmission potentials that are smaller than those required to generate the same final epidemic size on disassortatively-mixed populations (Fig. 4a). However, for large final epidemic sizes, the opposite is true. In this case, for disassortatively-mixed populations (Fig. 4a), large final epidemic sizes are obtained upon using transmission potentials that are smaller than those required to obtain the same final epidemic size on assortatively-mixed populations (Fig. 4b).

The final epidemic size is now considered based on numerical solutions of the differential equations given by Eqs. (12)–(15). In Fig. 5a, the final epidemic size is plotted as a function of the per contact transmission rate τ for a range of different levels of mixing. We note that initially the disease is seeded in both poorly and highly connected groups (i_L and i_H). This is especially important when the system consists of two noninteracting subpopulations ($M = 1$). This situation is captured by the step-like final epidemic size curve. For small values of τ, this curve exclusively represents the highly connected subpopulation where $R_0^H = \rho_H > 1$ while $R_0^L = \rho_L < 1$. The value of τ for which $\rho_L = 1$ corresponds to the step-like transition. When τ is large enough, ρ_L becomes greater than one and the disease can spread in both subpopulations. The mixing pattern of the population has considerable effect on whether an epidemic outbreak will happen. Assortatively

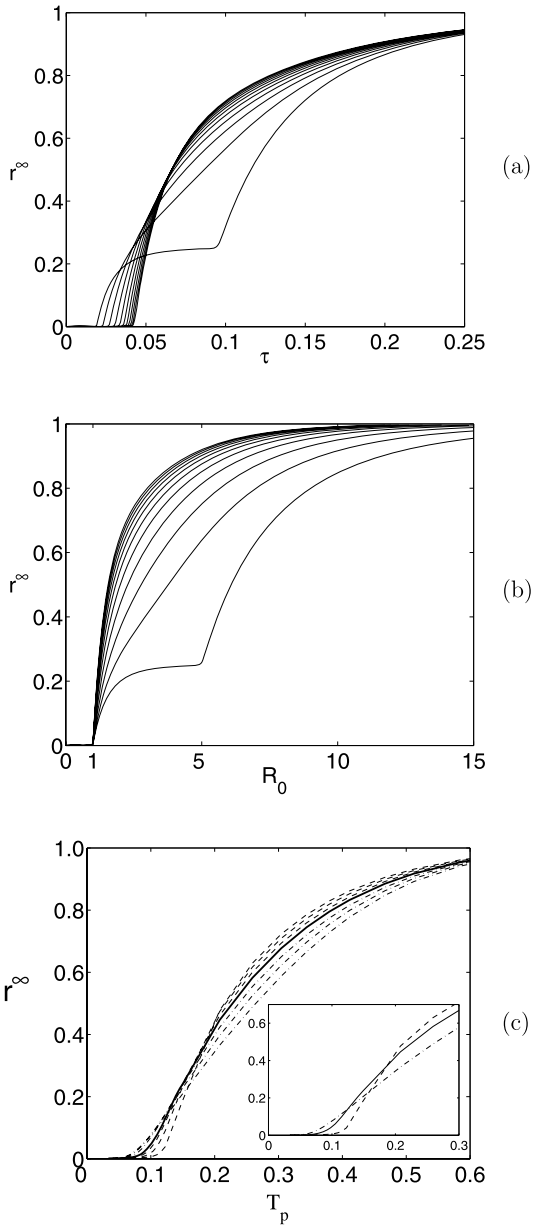

Fig. 5 Final epidemic size as a function of the per contact transmission rate τ (a) and basic reproduction number R_0 (b) for $M = -1.0, -0.8, \ldots, 1.0$ from right to left (a) and left to right (b). The case of $a = b$ is considered here with $k = 3$, $\sigma = 5$, $f = 0.25$ and a value of $\gamma \simeq 0.286$. (c) Final epidemic size based on individual network simulation for disassortatively mixed (dashed line, $M \approx -0.2, -0.10, -0.05$), random (solid line, $M \approx 0$) and assortatively mixed (dot-dashed line, $M \approx 0.05, 0.10, 0.20$) networks versus probability of transmission ($T_p = \tau/(\tau + \gamma)$). Simulation based on networks with $N = 10000$ nodes and degree distribution $p(l) \sim l^{-\alpha} e^{-l/L}$ with $L = 100$ and $\alpha = 2.5$.

mixed populations are more prone to epidemic outbreaks and disease spread requires a lower infectious rate than on disassortatively mixed populations. The final epidemic size as a function of the basic reproduction number (Fig. 5b) for different levels of mixing illustrates that values of $R_0 > 1$ result in larger epidemics on disassortatively mixed populations. Hence, assortatively mixed populations are more prone to epidemics, but experience smaller epidemics than disassortatively mixed populations. In Fig. 5c, we show plots of the final epidemic size based on the simulation of an SIR model on networks with $N = 10000$ nodes and with different levels of mixing (see Kiss et al., 2008). Networks are generated based on a link rewiring algorithm (Newman, 2003) that conserves the degree distribution, but allows to tune the level of mixing within the network. At least qualitatively, there is good agreement between the results based on the simple model and simulation results. Results from both models show a faster initial spread if mixing is assortative, and for high enough transmission rates, a higher final epidemic size if mixing is disassortative.

In the limit of $\tau \to \infty$, it is possible to derive analytical results for the final epidemic size. This analysis is motivated by the tendency of faster convergence of the final epidemic size to full population size for increasing values of M (Figs. 4 and 5). Based on Eqs. (28) and (29), the final epidemic size $r(\infty) = r_L(\infty) + r_H(\infty)$ is determined by the original ODE system through

$$r_L(\infty) = (1-f)\big[1 - \exp\big(-\tau\big(a_{LL}r_L(\infty) + a_{HL}r_H(\infty)\big)\big)\big], \tag{35}$$

$$r_H(\infty) = f\big[1 - \exp\big(-\tau\big(a_{LH}r_L(\infty) + a_{HH}r_H(\infty)\big)\big)\big]. \tag{36}$$

These implicit equations determine the value of $r_L(\infty)$ and $r_H(\infty)$. The nonlinear system given by Eqs. (35) and (36) can be solved numerically by using Newton iteration. Therefore, the final epidemic size can be computed without solving the differential equations. However, the nonlinear system cannot be solved explicitly, thus the qualitative properties of the final epidemic size curves in Fig. 5 are difficult to explain by using Eqs. (35) and (36). In the limit of large τ, however, it is possible to derive explicit approximating formulas for $r_L(\infty)$ and $r_H(\infty)$. Let us introduce two new variables X, Y instead of $r_L(\infty)$ and $r_H(\infty)$ such that

$$r_L(\infty) = (1-f)\big[1 - X\exp(-\rho_L)\big], \qquad r_H(\infty) = f\big[1 - Y\exp(-\rho_H)\big]. \tag{37}$$

Substituting these expressions into Eqs. (35) and (36) and upon using Eq. (16) and recalling that $\rho_L = \frac{\tau k}{\gamma}$ and $\rho_H = \frac{\tau\sigma k}{\gamma}$, the following equations are obtained

$$X = c_L \exp\left(\rho_L \frac{b}{1-a}\exp(-\rho_L)X + \rho_L \frac{1-a-b}{1-a}\exp(-\rho_H)Y\right), \tag{38}$$

$$Y = c_H \exp\left(\rho_H \frac{1-a-b}{1-b}\exp(-\rho_L)X + \rho_H \frac{a}{1-b}\exp(-\rho_H)Y\right), \tag{39}$$

where $c_L = \exp(\rho_L(1 - P(L|L) - P(H|L)))$ and $c_H = \exp(\rho_H(1 - P(L|H) - P(H|H)))$. From Eqs. (2) and (3) follows that $1 - P(L|L) - P(H|L) = 1 - P(L|H) - P(H|H) = 0$, and hence $c_L = c_H = 1$. Using this in Eqs. (38) and (39), the following nonlinear system for X and Y is obtained

$$X = \exp(b_{11}X + b_{12}Y), \qquad Y = \exp(b_{21}X + b_{22}Y), \tag{40}$$

where

$$b_{11} = \rho_L \frac{b}{1-a} \exp(-\rho_L), \qquad b_{12} = \rho_L \frac{1-a-b}{1-a} \exp(-\rho_H), \tag{41}$$

$$b_{21} = \rho_H \frac{1-a-b}{1-b} \exp(-\rho_L), \qquad b_{22} = \rho_H \frac{a}{1-b} \exp(-\rho_H). \tag{42}$$

When $\tau \to \infty$, the transmission potentials ρ_L and ρ_H tend to infinity. These new variables, however, allow to relocate this singularity to $b_{ij} \to 0$ as $\tau \to \infty$. Hence, we look for the solution (X, Y) of system (40) in the form of a power series around zero. Thus, let us substitute the following expansions:

$$X = 1 + c_{11}b_{11} + c_{12}b_{12} + c_{21}b_{21} + c_{22}b_{22} + h.o.t.,$$
$$Y = 1 + d_{11}b_{11} + d_{12}b_{12} + d_{21}b_{21} + d_{22}b_{22} + h.o.t.$$

into equations given by (40). Using the expansion of the exponential function and equating the coefficients of b_{ij}, we obtain

$$c_{11} = 1, \quad c_{12} = 1, \quad c_{21} = 0, \quad c_{22} = 0,$$
$$d_{11} = 0, \quad d_{12} = 0, \quad d_{21} = 1, \quad d_{22} = 1.$$

Hence, the first order approximation of the solutions of system (40) is

$$X = 1 + b_{11} + b_{12}, \qquad Y = 1 + b_{21} + b_{22}.$$

Substituting these expansions into Eq. (37) and using Eqs. (41) and (42), the following approximating formulas are obtained:

$$r_L(\infty) = (1-f)\Bigg[1 - \exp(-\rho_L) - \rho_L \frac{b}{1-a} \exp(-2\rho_L) - \rho_L \frac{1-a-b}{1-a} \exp(-\rho_L - \rho_H)\Bigg],$$

$$r_H(\infty) = f\Bigg[1 - \exp(-\rho_H) - \rho_H \frac{a}{1-b} \exp(-2\rho_H) - \rho_H \frac{1-a-b}{1-b} \exp(-\rho_L - \rho_H)\Bigg].$$

Adding these two equations and using $\rho_H = \sigma\rho_L$, we note that for large τ and $\sigma \geq 3$ the term $\exp(-\rho_H)$ is negligible with respect to $\exp(-\rho_L)$. Therefore, we obtain

$$r(\infty) = 1 - (1-f)\exp(-\rho_L) - (1-f)\rho_L \frac{b}{1-a} \exp(-2\rho_L) + h.o.t. \tag{43}$$

where $h.o.t.$ stands for terms smaller than $\exp(-2\rho_L)$.

The approximation given in Eq. (43) is in good agreement with the exact solution of the final epidemic size obtained using Newton's method (Fig. 6a). When $a = b$ (as in Fig. 5)

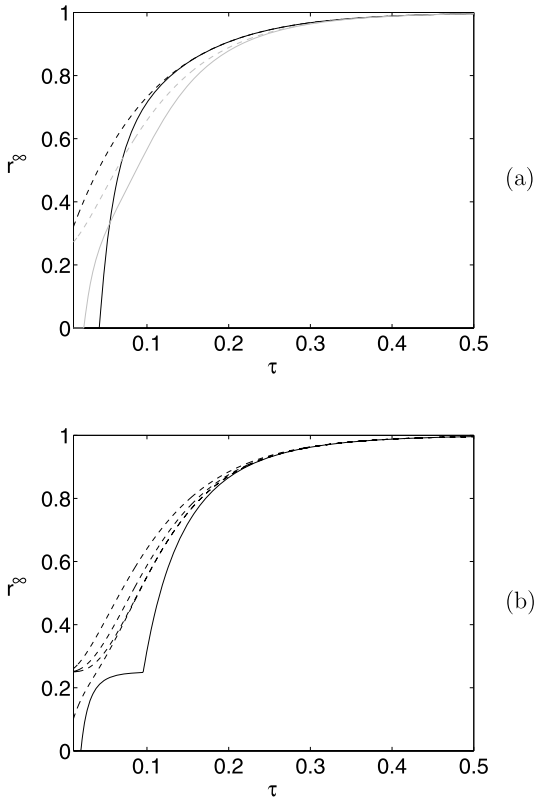

Fig. 6 (a) Illustration of the approximating formula given by Eq. (43) (dashed lines) versus numerical solution (continuous lines) for $M = -0.8$ (black lines) and $M = 0.8$ (grey lines) (b) Illustration of the approximating formula given by Eq. (44) (dashed lines) with increasing number of terms versus numerical solution (continuous line) for $M = 1.0$. For both (a) and (b), $k = 3$, $\sigma = 5$, $f = 0.25$, and $\gamma \simeq 0.286$.

from Eq. (43), we observe that $r(\infty)$ depends monotonically on M. This is justified by noting that in this case

$$-\frac{b}{1-a} = \frac{M+1}{M-3} = 1 + \frac{4}{M-3}$$

and this is a decreasing function of $M \in [-1, 1]$. Hence, the approximation is in good agreement with the numerical results presented in Fig. 5 and shows that the final epidemic size approaches one more rapidly when the population is disassortatively mixed.

In a similar way, higher order approximations can be derived for $r(\infty)$. As an example, we show a fourth order one in the case of $a + b = 1$ (this is the case when the approximations take relatively simple forms). For $\sigma \geq 5$, the approximation is

$$r(\infty) = 1 - cD - cD^2 - \frac{3}{2}cD^3 - \frac{8}{3}cD^4 + h.o.t. \tag{44}$$

where $c = (1 - f)/\rho_L$, $D = \rho_L \exp(-\rho_L)$ and *h.o.t.* stands for terms smaller than $\exp(-4\rho_L)$. The higher order formula give better approximations of $r(\infty)$ for smaller values of τ and this is illustrated in Fig. 6b where the approximation is plotted by increasing the number of terms in the approximation from one to four.

4. Discussion

Starting from a network representation of the contact pattern within a population, we have derived and analyzed a compartmental model that can capture preferential mixing. Such simple models provide a convenient and analytically tractable way to capture properties of the contact network that can have significant implications for disease transmission and control. We considered the simplest division of the population into poorly and highly connected individuals. However, our model can be easily extended to account for a more heterogeneous population provided that the connectivity correlations are also extended is some convenient way. The present model is flexible in that if connectivity correlations can be measured or estimated, then the model can be formulated in terms of real data and can be used to guide analyses concerning the potential for an epidemic outbreak.

The result derived from this simple model are in good qualitative agreement with results obtained from models used in the context of STDs (Gupta et al., 1989; Anderson et al., 1990; Kretzschmar et al., 1996; Ghani et al., 1997). The model results accurately show that assortatively mixed populations are more prone to epidemic outbreaks when compared to disassortatively mixed populations and that the value of R_0 is considerably higher when mixing within the population is assortative. However, the final epidemic size is smaller in assortatively mixed populations than in disassortatively mixed ones. This shows that depending on the mixing within the population, high values of R_0 can lead to epidemics that spread quickly but with final epidemic sizes that are small (i.e., assortatively mixed) and small values of R_0 can result in epidemics with a slow timescale, but infecting a considerable proportion of the population (i.e., disassortatively mixed). Hence, during an epidemic, an estimate of R_0 with no further information about the population contact structure cannot provide an accurate prediction about the outcome of an epidemic. For example, consider the case of assortative mixing when the proportion of highly connected individuals (f) varies. In Fig. 7, based on numerically solving Eqs. (35) and (36) by using Newton iteration, the final epidemic size is given as a function of R_0 for different values of f. From Fig. 7 follows that the presence of a highly connected core group, even if very small, leads to large values of R_0 that can only generate smaller and smaller final epidemic sizes as f decreases (Diekmann and Heesterbeek, 2000). For high levels of assortative mixing when the noninteracting subpopulation regime is approached and for f close to one, the final epidemic size approaches the result that is obtained from the standard SIR model.

By using individual-based stochastic network simulations and purely based on numerical results, Kiss et al. (2008) has obtained results that are qualitatively equivalent (Fig. 5). However, here analytical results allow us to identify a more precise relation between the mixing pattern and epidemic characteristics. Our results show that population contact structure and disease dynamics can interact in nontrivial ways and must be considered concurrently. The present model could also provide a basis for incorporating epidemic control measures (e.g., vaccination) with the aim of investigating the implications of the

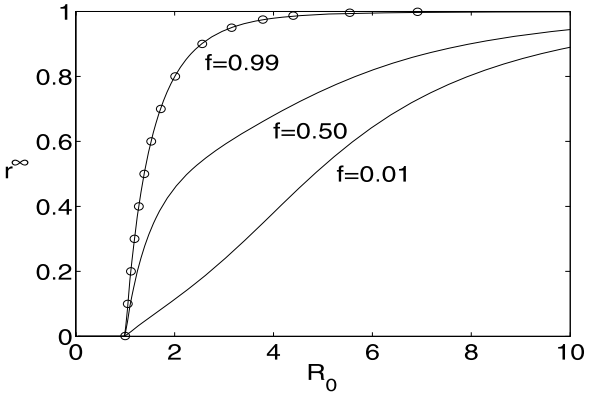

Fig. 7 Final epidemic size as a function of R_0 for different values of f. Open circles (○) denote the final epidemic size corresponding to the standard SIR model. In this case, $M = 0.8$, $k = 3$, $\sigma = 5$, and $\gamma \simeq 0.286$.

population contact structure properties for the effectiveness of various epidemic control measures.

Acknowledgement

The authors thank the reviewers for their careful consideration of the paper and their helpful comments, and O. Diekmann for useful discussions.

References

Anderson, R.M., Garnett, G., 2000. Mathematical models of the transmission and control of sexually transmitted disease. Sex. Transm. Dis. 27, 636–643.

Anderson, R.M., Gupta, S., Ng, W., 1990. The significance of sexual partner contact networks for the transmission of HIV. J. AIDS 3, 417–429.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, London.

Barthélemy, M., Barrat, A., Pastoras-Satorras, R., Vespignani, A., 2005. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. J. Theor. Biol. 235, 257–288.

Boguñá, M., Pastor-Satorras, R., Vespignani, A., 2003a. Absence of epidemic threshold in scale-free networks with degree-correlations. Phys. Rev. Lett. 90, 028701.

Boguñá, M., Pastor-Satorras, R., Vespignani, A., 2003b. Epidemics spreading in complex networks with degree correlations. In: Pastor-Satorras, R., Rubi, J.M., Guilera, A.D. (Eds.), Statistical Mechanics of Complex Networks, Lecture Notes in Physics, vol. 625, pp. 127–147. Springer, Berlin,

Catania, J.A., Coates, T.J., Kegels, S., Fullilove, M.T., 1992. Condom use in multi-ethnic neighborhoods of San Francisco: the population-based AMEN (AIDS in Multi- Ethnic Neighborhoods) study. Am. J. Publ. Health 82, 284–287.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley, New York.

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.

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. doi:10.1126/science.1061020.
Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.S., 2005. Strategies for containing an emerging influenza pandemic in South East Asia. Nature 437, 209–214. doi:10.1038/nature04017.
Ghani, A., Swinton, J., Garnett, G., 1997. The role of sexual partnership networks in the epidemiology of gonorrhea. Sex. Transm. Infect. 24, 45–56.
Green, D.M., Kiss, I.Z., Kao, R.R., 2006. Modelling the initial spread of foot-and-mouth disease through animal movements. Proc. R. Soc. B 273, 2729–2735. doi:10.1098/rspb.2006.3648.
Gupta, S., Anderson, R.M., May, R.M., 1989. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 3, 807–818.
Hethcote, H.W., Yorke, J.A., 1984. Gonorrhea Transmission Dynamics and Control. Lecture Notes in Biomathematics, vol. 56. Springer, New York.
Hufnagel, L., Brockmann, D., Geisel, T., 2004. Forecast and control of epidemics in a globalized world. Proc. Natl. Acad. Sci. USA 101, 15 124–15 129. doi:10.1073/pnas.0308344101.
Jones, H.J., Handcock, M.S., 2003. An assessment of preferential attachment as a mechanism for human sexual network formation. Proc. R. Soc. B 270, 1123–1128. doi:10.1098/rspb.2003.2369.
Kao, R.R., 2006a. Evolution of pathogens towards low R_0 in heterogeneous populations. J. Theor. Biol. 242, 634–642.
Kao, R.R., Danon, L., Green, D.M., Kiss, I.Z., 2006b. 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., et al., 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294, 813–817. doi:10.1126/science.1065973.
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.
Kiss, I.Z., Green, D.M., Kao, R.R., 2006a. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. J. R. Soc. Interface 3, 669–677. doi:10.1098/rsif.2006.0129.
Kiss, I.Z., Green, D.M., Kao, R.R., 2006b. The effect of network heterogeneity and multiple routes of transmission on final epidemic size. Math. Biosci. 203, 124–136. doi:10.1016/j.mbs.2006.03.002.
Kiss, I.Z., Green, D.M., Kao, R.R., 2008. The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. J. R. Soc. Interface 5, 791–799. doi:10.1098/rsif.2007.1272.
Kretzschmar, M., van Duynhoven, Y.T.H.P., Severijnen, A.J., 1996. Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations. Am. J. Epidemiol. 144, 306–317.
Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y., 2001. The web of human sexual networks. Nature 411, 907–908. doi:10.1038/35082140.
Lipsitch, M., et al., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966–1970. doi:10.1126/science.1086616.
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. doi:10.1016/j.jtbi.2004.07.026.
Moreno, Y., Gómez, J.B., Pacheco, A.F., 2003. Epidemic incidence in correlated complex networks. Phys. Rev. E 68, 035103(R).
Newman, M.E.J., 2002. Assortative mixing in networks. Phys. Rev. E 89, 208701.
Newman, M.E.J., 2003. Mixing patterns in networks. Phys. Rev. E 67, 026126.
Pastor-Satorras, R., Vazquez, A., Vespignani, A., 2001. Dynamical and correlation properties of the internet. Phys. Rev. Lett. 87, 0258701. doi:10.1103/PhysRevLett.87.258701.
van den Driessche, P., Watmough, J.J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180, 29–48.