



**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 number  $R_0$ . In this paper a simple differential equation model that accounts for connectivity correlations based on the number of contacts is given. The implications of disassortative and assortative mixing patterns for the epidemic threshold, epidemic dynamics and final epidemic size are determined by a combination of numerical and analytical approaches. In assortatively mixed populations, highly connected individuals become infected at a faster rate than poorly connected individuals and these are more difficult to infect due to the reduced number of contacts. Assortatively mixed populations are more prone to disease outbreaks with higher values of the basic reproduction number  $R_0$  for the same transmission rates but for sufficiently large transmission rates, experience smaller epidemics than disassortatively mixed populations.

**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 population data, the collection of which has been partially driven by the epidemics of HIV/AIDS (Liljeros et al., 2001; Jones & Handcock, 2003; Anderson et al., 1990; Catania et al., 1992), SARS (Lipsitch et al., 2001; 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; Green et al., 2006; Kao et al., 2006a; Kiss et al., 2006a), 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, spatial structure and connectivity correlations.

Network models where epidemic units (e.g. individuals, households, farming premises) are represented by nodes and potentially infectious contacts by links between the nodes are often used to accommodate contact heterogeneity and connectivity correlations. However, differential equation-based models can also be adapted to capture such properties and are amenable to a variety of powerful analytical tools. For example, May and Lloyd (2001) and Kiss et al. (2006b) used differential equation-based models to investigate the role of heterogeneity in the distribution of contact frequencies. Here we extend these approaches to account for connectivity correlations and investigate the implications of this property for infectious disease transmission.

In the network formulation, random or proportionate mixing means that the probability of any two nodes being connected is proportional to the product of their degree (i.e. number of contacts). However, where “like attaches to like” (e.g. for sexually transmitted infections or *STIs*, Anderson et al., 1990; Catania et al., 1992) or where “opposites attract” (e.g. livestock trading among farms and markets in the UK, Kiss et al., 2006a; correlation properties of the Internet, Pastor-Satorras et al., 2001) this is not the case. Empirical evidence has led to numerous network- and differential equation-based models (Boguñá et al., 2003a,b; 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 node degree variance grow rapidly, and in the limiting case of infinite variance, instantaneously, independently of the mixing pattern (Boguñá et al., 2003a).

These previous studies have derived bounds for the epidemic threshold in the case of infinite variance and in the presence of connectivity correlations. However, the implications for epidemic dynamics, and an analytical derivation of the epidemic threshold and final epidemic size have not been considered. Here we develop a simple differential equation model that accounts for connectivity correlations based on the number of contacts that individuals have, and analyze in detail implications for the

69 epidemic threshold, epidemic dynamics and final epidemic size.

## 70 2 Model

### 71 2.1 Population Mixing model

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

$$\langle k \rangle = \frac{kN_L + \sigma kN_H}{N_L + N_H}, \quad (1)$$

79 and upon using  $f = \frac{1}{\sigma-1}$  the average number of connection per individual can be  
 80 constrained so that  $\langle k \rangle = 2k$ . Therefore, the average number of connections per  
 81 individual over the whole population is twice the number of connections that poorly  
 82 connected individuals have. If contact would occur at random then the probability of  
 83 connection between two individuals is proportional to the product of their degree (i.e.  
 84 proportionate mixing). However, we wish to model the situation when individuals in  
 85 one particular group may or may not preferentially mix or contact individuals of the  
 86 same type. To accommodate this preferential mixing we assume that a proportion  
 87  $0 \leq a \leq 1$  of all contacts within the population occur between highly connected  
 88 individuals. Similarly, let  $0 \leq b \leq 1$  represent the proportion of contacts between  
 89 individuals that are less well connected. Therefore,  $0 \leq 1 - a - b \leq 1$  ( $\Rightarrow a +$   
 90  $b \leq 1$ ) represents the proportion of contacts between poorly and highly connected  
 91 individuals. This leads to defining the following connection correlations within the  
 92 population

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

93 and

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

94 The conditional probabilities defined in Eqs. (2) and (3) are bounded below by  
 95 zero and from above by one and they satisfy  $P(L|L) + P(H|L) = 1$  and  $P(L|H) +$   
 96  $P(H|H) = 1$ . This means that all links starting at a highly connected individual will  
 97 connect to poorly connected individuals with probability  $P(L|H)$  and to highly con-  
 98 nected individuals with probability  $P(H|H)$ . Based on the connectivity correlations

we can define a mixing matrix that allows to measure the level and type of mixing within the population

$$\begin{matrix} & L & H & a_i \\ \begin{matrix} L \\ H \\ b_i \end{matrix} & \begin{pmatrix} b & \frac{1-a-b}{2} & \frac{1-a+b}{2} \\ \frac{1-a-b}{2} & a & \frac{1+a-b}{2} \\ \frac{1-a+b}{2} & \frac{1+a-b}{2} & 1 \end{pmatrix} & = E. \end{matrix} \quad (4)$$

The entries of the matrix ( $e_{ij}$ ) represent the proportion of connections/links within and between the two different sub-groups. The sum of the rows and columns are denoted by  $a_i$  and  $b_i$  respectively. To quantify the level of mixing within the population consider the assortativity coefficient (Newman, 2002, 2003) defined as

$$r = \frac{\sum_{i \in \{L,H\}} e_{ii} - \sum_{i \in \{L,H\}} a_i b_i}{1 - \sum_{i \in \{L,H\}} a_i b_i} = \frac{a + b - \left(\frac{1-a+b}{2}\right)^2 - \left(\frac{1+a-b}{2}\right)^2}{1 - \left(\frac{1-a+b}{2}\right)^2 - \left(\frac{1+a-b}{2}\right)^2} \in [-1, 1]. \quad (5)$$

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

## 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  $a + b = 1$  follows that  $r = 1$ , and in this case the population consists of two non-interacting sub-populations. Therefore, transmission is not possible between poorly and highly connected individuals. Based on these

126 assumptions we have the following system of differential equations:

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

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

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

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

$$\frac{dR_L}{dt} = \gamma i_L, \quad (10)$$

$$\frac{dR_H}{dt} = \gamma i_H, \quad (11)$$

127 where  $\tau$  is the per contact rate of transmission and  $\gamma$  is the recovery rate. The right  
 128 hand side in Eq. (6) describes the creation of new infections and is proportional to  
 129 the transmission rate  $\tau$ , the number of contacts  $k$  of poorly connected susceptible  
 130 individuals, the number of susceptible individuals with  $k$  connections, and the prob-  
 131 ability that any given neighbour of a susceptible individual with  $k$  connections is  
 132 infected. The formulation for such type of heterogeneous contact follows from that of  
 133 Anderson & May (1991) for sexually transmitted diseases. A similar approach is used  
 134 when modelling correlated complex networks (Boguñá et al., 2003a,b; Barthélemy  
 135 et al., 2005). In this present model, following May and Lloyd (2001), it is assumed  
 136 that contacts switch at a rate that is much faster than the rate at which disease  
 137 spreads, and thus infection does not result in the loss of one susceptible neighbour  
 138 upon becoming infectious. Eqs. (6) to (11) are non-dimensionalised using  $N$  to scale  
 139 all variables and  $1/\gamma$  to scale time. We note that it is sufficient to consider the first  
 140 four equations since the values of  $r_L$  and  $r_H$  are determined by the values of the  
 141 other four variables. Hence, the reduced system is given by

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

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

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

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

142 where

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

### 143 3 Results

#### 144 3.1 The basic reproduction number $R_0$

145 Following van den Driessche and Watmough (2002) we note that only compartments  
 146  $i_L$  and  $i_H$  are involved in the calculation of  $R_0$ . At the disease-free equilibrium  
 147  $(s_L, s_H, i_L, i_H) = (1 - f, f, 0, 0)$  matrices  $F$  and  $V$  are given by

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

148 The basic reproduction number  $R_0$  is defined as the leading eigenvalue of the next gen-  
 149 eration matrix  $FV^{-1}$ . Solving the resulting quadratic equation the leading eigenvalue  
 150 and hence  $R_0$  is given by

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

151 High values of  $\sigma$  denote high levels of heterogeneity with a considerable difference  
 152 between the number of connections of individuals in the two sub-populations. Higher  
 153 number of connections lead to a higher number of individuals being infected by an  
 154 index case and therefore to a higher value of  $R_0$ . This is illustrated in Fig. 2 where  
 155 contour plots of  $R_0$  are given for increasing values of  $\sigma$  while keeping all disease  
 156 spread and disease progression related parameters constant. Fig. 2 also illustrates  
 157 how higher levels of assortativity (i.e. in Fig. 2 the region close to  $a + b = 1$ ) lead  
 158 to higher values of  $R_0$ . Therefore, for the same disease, an outbreak is more likely  
 159 to happen if the population is assortatively mixed. At the individual level highly  
 160 connected individuals preferentially connect to other highly connected individuals  
 161 and this leads to a high number of secondary infections generated by an initial index  
 162 case.

163 The transmission potentials within the individual sub-populations can be defined  
 164 as  $\rho_L = \frac{\tau k}{\gamma}$  and  $\rho_H = \frac{\tau \sigma k}{\gamma}$ . These represent the number of secondary infections  
 165 produced by a single infectious individual during its infectious period when introduced  
 166 into a fully susceptible sub-population. The basic reproduction number  $R_0$  has a  
 167 similar definition but the index case is weighted according to the probability of it itself  
 168 becoming infected. In the case of non-interacting sub-populations the transmission  
 169 potentials are equivalent to  $R_0$  within the sub-population. The basic reproduction  
 170 number  $R_0$  in terms of the transmission potentials can be rewritten to give

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

171 The limit of  $r = -1$  represents a disassortatively mixed population meaning that  
 172 there are no within sub-population connections. Therefore, ( $P(L|L) = P(H|H) =$   
 173  $0, P(L|H) = P(H|L) = 1$ ) and  $R_0$  in this case is given by

$$R_0 = \sqrt{\rho_L \rho_H}. \quad (20)$$

174 This represents the geometric mean of the individual basic reproduction numbers  
 175 corresponding to individual sub-populations. Hence, this stands for a complete cy-  
 176 cle of infection from poorly to highly connected individuals and then back again to  
 177 poorly connected individuals (Diekmann & Heesterbeek, 2000). Such a cycle results  
 178 in  $\rho_L \rho_H$  new infections. If  $r = 1$ , the population is divided into two distinct and  
 179 non-interacting sub-populations ( $P(L|H) = P(H|L) = 0, P(L|L) = P(H|H) = 1$ )  
 180 and the basic reproduction number is given by

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

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

### 183 3.2 Epidemic dynamics

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

### 202 3.3 Final epidemic size

203 The severity of the disease is given by the final epidemic size  $r(\infty) = r^\infty = r_L(\infty) +$   
 204  $r_H(\infty) = r_L^\infty + r_H^\infty$  and we derive analytical results in the form of a family of parametric  
 205 equations that link the final epidemic size and the two transmission potentials. For  
 206 ease of notation Eqs. (12)-(15) are rewritten to give

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

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

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

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

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

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

210 where

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

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

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

213 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  
 214  $m \in \{L, H\}$ . Considering the limit of  $t \rightarrow \infty$  in Eq. (26) it follows that

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

215 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  
 216 the limit of  $t \rightarrow \infty$  we obtain

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

217 and

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

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

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

219 Combining Eq. (30) and the re-arranged version of Eq. (31) and upon using the  
220 notations given in Eq. (16) the following parametric equations are obtained

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

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

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

221 In Fig. 4, based on Eqs. (32) to (34), plots of the final epidemic size ( $r^\infty$ ) are given  
222 for a range of transmission potentials for disassortatively- and assortatively-mixed  
223 populations. We separately consider the case of small and large final epidemic size  
224 and discuss the dependence of  $r^\infty$  on the values of the transmission potentials. For  
225 assortatively-mixed populations (Fig. 4b) an initial perturbation results in small  
226 final epidemic sizes for transmission potentials that are smaller than those required  
227 to generate the same final epidemic size on disassortatively-mixed populations (Fig.  
228 4a). However, for large final epidemic sizes the opposite is true. In this case, for  
229 disassortatively-mixed populations (Fig. 4a), large final epidemic sizes are obtained  
230 upon using transmission potentials that are smaller than those required to obtain the  
231 same final epidemic size on assortatively-mixed populations (Fig. 4b).

232 The final epidemic size is now considered based on numerical solutions of the  
233 differential equations given by Eqs. (12) to (15). In Fig. 5a the final epidemic size  
234 is plotted as a function of the per contact transmission rate  $\tau$  for a range of different  
235 levels of mixing. The case of non-interacting sub-populations ( $r = -1.0$ ) is captured  
236 by the step-like final epidemic size curve. For small values of  $\tau$  this curve exclusively  
237 represents the highly connected sub-population where  $R_0 = \rho_H > 1$  while  $R_0 = \rho_L <$   
238 1. When  $\tau$  is large enough,  $\rho_L$  becomes greater than one and epidemic outbreaks  
239 within the poorly connected sub-population are also possible. The mixing pattern of  
240 the population has considerable effect on whether an epidemic outbreak will happen.  
241 Assortatively mixed populations are more prone to epidemic outbreaks and disease  
242 spread requires a lower infectious rate than on disassortatively mixed populations.

243 The final epidemic size as a function of the basic reproduction number (Fig 5b) for  
 244 different levels of mixing illustrates that values of  $R_0 > 1$  result in larger epidemics  
 245 on disassortatively mixed populations. Hence, assortatively mixed populations are  
 246 more prone to epidemics but experience smaller epidemics than disassortatively mixed  
 247 populations.

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

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

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

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

$$r_L(\infty) = (1 - f) [1 - X \exp(-\rho_L)], \quad r_H(\infty) = f [1 - Y \exp(-\rho_H)]. \quad (37)$$

261 Substituting these expressions into Eqs. (35) and (36) the following nonlinear system  
 262 for  $X$  and  $Y$  is obtained

$$X = \exp(b_{11}X + b_{12}Y), \quad Y = \exp(b_{21}X + b_{22}Y), \quad (38)$$

263 where

$$b_{11} = \rho_L \frac{b}{1 - a} \exp(-\rho_L), \quad b_{12} = \rho_L \frac{1 - a - b}{1 - a} \exp(-\rho_H), \quad (39)$$

$$b_{21} = \rho_H \frac{1 - a - b}{1 - b} \exp(-\rho_L), \quad b_{22} = \rho_H \frac{a}{1 - b} \exp(-\rho_H) \quad (40)$$

When  $\tau \rightarrow \infty$ , the transmission potentials  $\rho_L$  and  $\rho_H$  tend to infinity. These new  
 variables however, allow to relocate this singularity to  $b_{ij} \rightarrow 0$  as  $\tau \rightarrow \infty$ . Hence we  
 look for the solution  $(X, Y)$  of system (38) 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 (38). Using the expansion of the exponential function and equating the coefficients of  $b_{ij}$  we obtain

$$\begin{aligned} c_{11} &= 1, & c_{12} &= 1, & c_{21} &= 0, & c_{22} &= 0, \\ d_{11} &= 0, & d_{12} &= 0, & d_{21} &= 1, & d_{22} &= 1. \end{aligned}$$

Hence the first order approximation of the solutions of system (38) is

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

Substituting these expansions into Eq. (37) and using Eqs. (39) and (40) the following approximating formulas are obtained

$$\begin{aligned} r_L(\infty) &= (1-f) \left[ 1 - \exp(-\rho_L) - \rho_L \frac{b}{1-a} \exp(-2\rho_L) - \rho_L \frac{1-a-b}{1-a} \exp(-\rho_L - \rho_H) \right], \\ r_H(\infty) &= f \left[ 1 - \exp(-\rho_H) - \rho_H \frac{a}{1-b} \exp(-2\rho_H) - \rho_H \frac{1-a-b}{1-b} \exp(-\rho_L - \rho_H) \right]. \end{aligned}$$

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

$$r(\infty) = 1 - (1-f) \exp(-\rho_L) - (1-f) \rho_L \frac{b}{1-a} \exp(-2\rho_L) + h.o.t. \quad (41)$$

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

The approximation given in Eq. (41) 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) from Eq. (41) we observe that  $r(\infty)$  depends monotonically on  $r$ . This is justified by noting that in this case

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

267 and this is an decreasing function of  $r$ . Hence, the approximation is in good agreement  
268 with the numerical results presented in Fig. 5 and shows that the final epidemic size  
269 approaches one more rapidly when the population is disassortatively mixed.

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

$$r(\infty) = 1 - cD - cD^2 - \frac{3}{2}cD^3 - \frac{8}{3}cD^4 + h.o.t. \quad (42)$$

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

## 277 4 Discussion

278 This simple model captures preferential mixing based on the number of contacts that  
279 individuals have. Such simple models provide a convenient and analytically tractable  
280 way to identify the implications of various population contact structure properties  
281 for disease transmission and control. The present study is motivated by empirical  
282 evidence such as the correlation properties of *STIs* and the Internet. We considered  
283 the simplest division of the population into poorly and highly connected individuals.  
284 However, our model can be easily extended to account for a more heterogeneous pop-  
285 ulation provided that connectivity probabilities are also extended in some convenient  
286 way. The present model is flexible in that if connectivity correlations can be measured  
287 or estimated then the model can be formulated in terms of real data and can be used  
288 to guide some useful analysis concerning the potential for an epidemic outbreak.

289 We have shown that assortatively mixed populations are more prone to epidemic  
290 outbreaks when compared to disassortatively mixed populations. The value of  $R_0$  was  
291 considerably higher when mixing within the population was assortative. Since in the  
292 assortatively mixed case like attaches to like, highly connected nodes are infected and  
293 depleted faster while many poorly connected nodes are still susceptible. This results  
294 in final epidemic sizes that are smaller on assortatively mixed populations than on  
295 disassortatively mixed ones. This shows that, depending on the mixing within the  
296 population, high values of  $R_0$  can lead to epidemics that spread quickly but with final  
297 epidemic sizes that are small (i.e. assortatively mixed) and small values of  $R_0$  can  
298 result in epidemics with a slow timescale but infecting a considerable proportion of  
299 the population (i.e. disassortatively mixed). Hence, during an epidemic, an estimate  
300 of  $R_0$  with no further information about the population contact structure cannot  
301 provide an accurate prediction regarding the outcome of an epidemic.

302 By using individual-based stochastic network simulations and purely based on nu-  
303 merical results, Kiss et al. (2008) obtained results that are qualitatively equivalent.  
304 However, here analytical results allow us to identify a more precise relation between  
305 the mixing pattern and epidemic characteristics. Our results show that population  
306 contact structure and disease dynamics can interact in non-trivial ways and must be  
307 considered concurrently. The present model could also provide a basis for incorpo-  
308 rating epidemic control measures (e.g. vaccination, contact tracing) with the aim  
309 to investigate the implications of the population contact structure properties for the  
310 effectiveness of various epidemic control measures.

## References

- 311
- 312 [1] Anderson, R. M. & May, R. M., 1991. *Infectious diseases of humans: dynamics*  
313 *and control*, Oxford University Press.
- 314 [2] Anderson, R. M., Gupta, S., Ng, W., 1990. The significance of sexual partner  
315 contact networks for the transmission of HIV. *J. AIDS* 3, 417429.
- 316 [3] Barthélemy, M., Barrat, A., Pastoras-Satorras, R., Vespignani A., 2005. Dynam-  
317 ical patterns of epidemic outbreaks in complex heterogeneous networks. *J. Theor.*  
318 *Biol.* 235, 257 - 288.
- 319 [4] Boguñá, M., Pastor-Satorras, R., Vespignani, A., 2003a. Absence of epidemic  
320 threshold in scale-free networks with degree-correlations. *Phys. Rev. Let.* 90,  
321 028701.
- 322 [5] Boguñá, M., Pastor-Satorras, R., Vespignani, A., 2003b. Epidemics spreading in  
323 complex networks with degree correlations. In "Statistical Mechanics of Complex  
324 Networks" (127 - 147), eds. Pastor-Satorras, R., Rubi. J. M. & Guilerá A.D.,  
325 Lecture Notes in Physics, vol. 625. Springer, Berlin.
- 326 [6] Catania, J. A., Coates, T. J., Kegels, S., Fullilove, M. T., 1992. Condom use  
327 in multi-ethnic neighborhoods of San Francisco: the population-based AMEN  
328 (AIDS in Multi- Ethnic Neighborhoods) study. *Am. J. Public Health* 82, 284287.
- 329 [7] Diekmann, O., Heesterbeek, J. A. P., 2000. *Mathematical epidemiology of infec-*  
330 *tious diseases: model building, analysis and interpretation*. John Wiley & Sons  
331 Ltd, Chichester, UK.
- 332 [8] Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A.,  
333 Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban  
334 social networks. *Nature* 429, 180184. (doi:10.1038/nature02541)
- 335 [9] Ferguson, N. M., Donnelly, C. A., Anderson, R. M., 2001. The foot-and mouth  
336 epidemic in Great Britain: pattern of spread and impact of interventions. *Science*  
337 292, 11551160. (doi:10.1126/science.1061020)
- 338 [10] Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S.,  
339 Meeyai, A., Iamsirithaworn, S., Burke, D. S., 2005. Strategies for contain-  
340 ing an emerging influenza pandemic in South East Asia. *Nature* 437, 209214.  
341 (doi:10.1038/nature04017)
- 342 [11] Green, D. M., Kiss, I. Z., Kao, R. R., 2006. Modelling the initial spread of foot-  
343 and-mouth disease through animal movements. *Proc. R. Soc. B* 273, 27292735.  
344 (doi:10.1098/rspb.2006.3648)

- 345 [12] Hufnagel, L., Brockmann, D., Geisel, T., 2004. Forecast and control of epi-  
346 demics in a globalized world. Proc. Natl Acad. Sci. USA 101, 15 12415 129.  
347 (doi:10.1073/pnas. 0308344101)
- 348 [13] Jones, H. J., Handcock, M. S., 2003. An assessment of preferential attachment as  
349 a mechanism for human sexual network formation. Proc. R. Soc. B 270, 11231128.  
350 (doi:10.1098/rspb.2003.2369)
- 351 [14] Kao, R. R., Danon, L., Green, D. M., Kiss, I. Z., 2006a. Demographic structure  
352 and pathogen dynamics on the network of livestock movements in Great Britain.  
353 Proc. R. Soc. B 273, 19992007. (doi:10.1098/rspb.2006.3505)
- 354 [15] Kao, R. R., 2006b. Evolution of pathogens towards low  $R_0$  in heterogeneous  
355 populations. J. Theor. Biol. 242 634642.
- 356 [16] Keeling, M. J. et al., 2001. Dynamics of the 2001 UK foot and mouth epi-  
357 demic: stochastic dispersal in a heterogeneous landscape. Science 294, 813817.  
358 (doi:10.1126/science. 1065973)
- 359 [17] Kiss, I. Z., Green, D. M., Kao, R. R., 2005. Disease contact tracing in random  
360 and clustered networks. Proc. R. Soc. B 272, 1407-1414.
- 361 [18] Kiss, I. Z., Green, D. M., Kao, R. R., 2006a. The network of sheep movements  
362 within Great Britain: network properties and their implications for infectious  
363 disease spread. J.R. Soc. Interface 3, 669677. (doi:10.1098/rsif.2006.0129)
- 364 [19] Kiss, I. Z., Green, D. M., Kao, R. R., 2006b. The effect of network heterogeneity  
365 and multiple routes of transmission on final epidemic size. Math. Biosci. 203, 124  
366 - 136. (doi:10.1016/j.mbs.2006.03.002)
- 367 [20] Kiss, I. Z., Green, D. M., Kao, R. R., 2008. The effect of network mixing patterns  
368 on epidemic dynamics and the efficacy of disease contact tracing. J. R. Soc.  
369 Interface 5, 791799. (doi:10.1098/rsif.2007.1272)
- 370 [21] Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., Aberg, Y., 2001.  
371 The web of human sexual networks. Nature 411, 907908. (doi:10.1038/35082140)
- 372 [22] Lipsitch, M. et al., 2003. Transmission dynamics and control of severe acute  
373 respiratory syndrome. Science 300, 19661970. (doi:10.1126/science.1086616)
- 374 [23] May, R. M., Lloyd, A. L., 2001. Infection dynamics on scale-free networks, Phys.  
375 Rev. E 64, 066112.

- 376 [24] Meyers, L. A., Pourbohloul, B., Newman, M. E. J., Skowronski, D. M., Brunham,  
 377 R. C., 2005. Network theory and SARS: predicting outbreak diversity. *J. Theor.*  
 378 *Biol.* 232, 7181. (doi:10.1016/j.jtbi.2004.07.026)
- 379 [25] Moreno, Y., Gómez, J. B., Pacheco, A. F., 2003. Epidemic incidence in correlated  
 380 complex networks. *Phys. Rev. E* 68, 035103(R).
- 381 [26] Newman, M. E. J., 2002. Assortative mixing in networks. *Phys. Rev. E* 65,  
 382 208701.
- 383 [27] Newman, M. E. J., 2003. Mixing patterns in networks. *Phys. Rev. E* 67, 026126.
- 384 [28] Pastor-Satorras, R., Vazquez, A., Vespignani, A., 2001. Dynamical and  
 385 correlation properties of the internet. *Phys. Rev. Lett.* 87, 0 258 701.  
 386 (doi:10.1103/PhysRevLett. 87.258701)
- 387 [29] van den Driessche, P., James Watmough, J., 2002. Reproduction numbers and  
 388 sub-threshold endemic equilibria for compartmental models of disease transmis-  
 389 sion, *Mathematical Biosciences* 180, 2948.

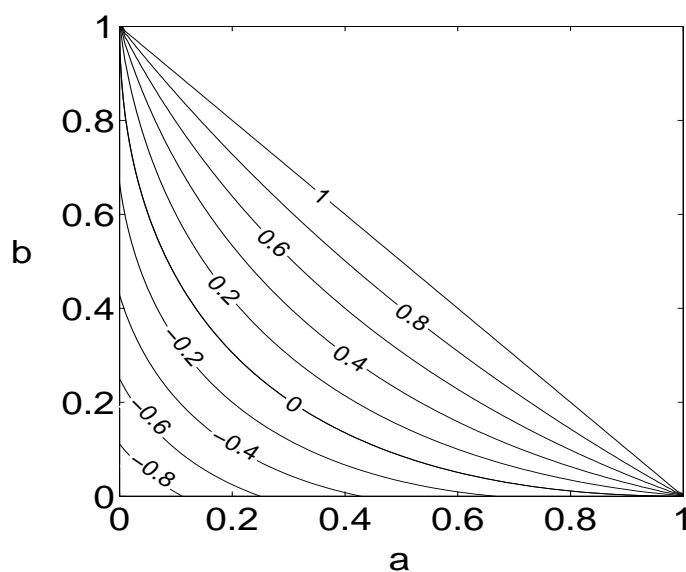


Figure 1: Illustration of  $r$ , defined in equation (5), as a function of  $a$  and  $b$  such that  $a + b \leq 1$ .

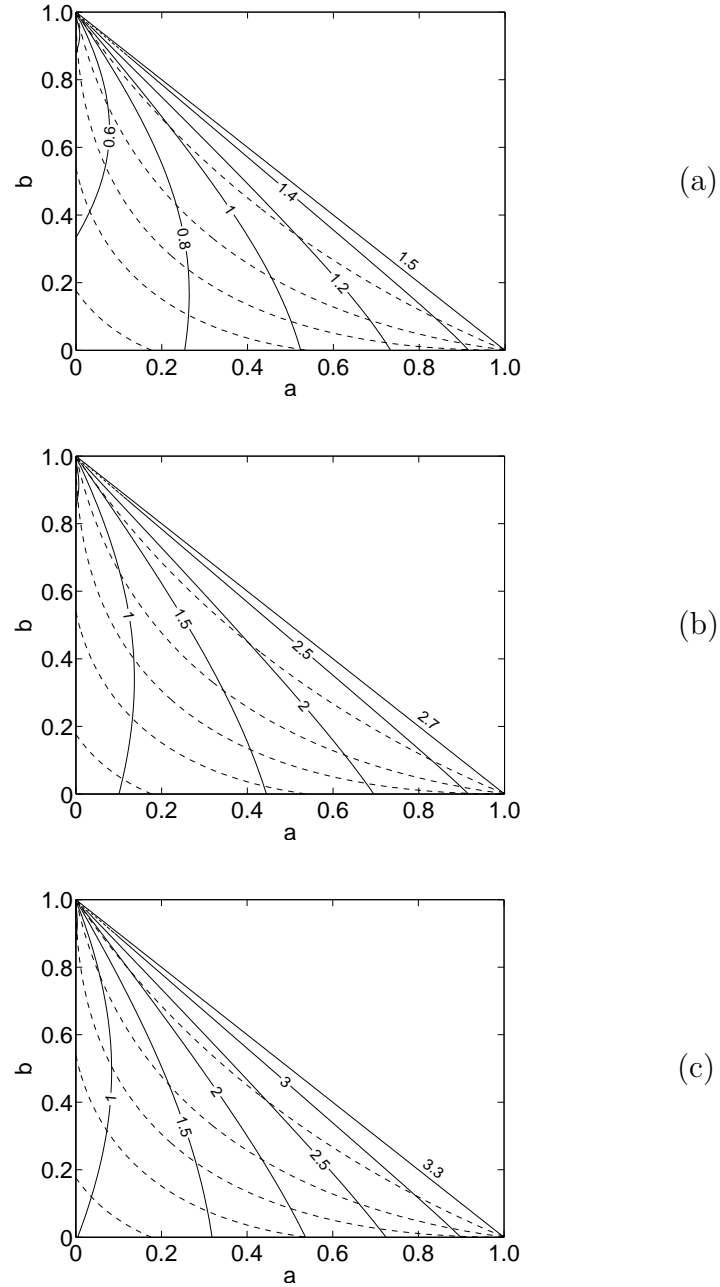


Figure 2: Based on equation (18),  $R_0$  plotted as a function of  $a$  and  $b$  for  $\sigma = 5, 9$ , and  $11$  in (a), (b) and (c) respectively. For all plots,  $\tau k/2\gamma = 0.15$ . Dotted lines from left to right show the various levels of mixing  $r = -0.7, -0.3, 0, 0.3, 0.7$ .

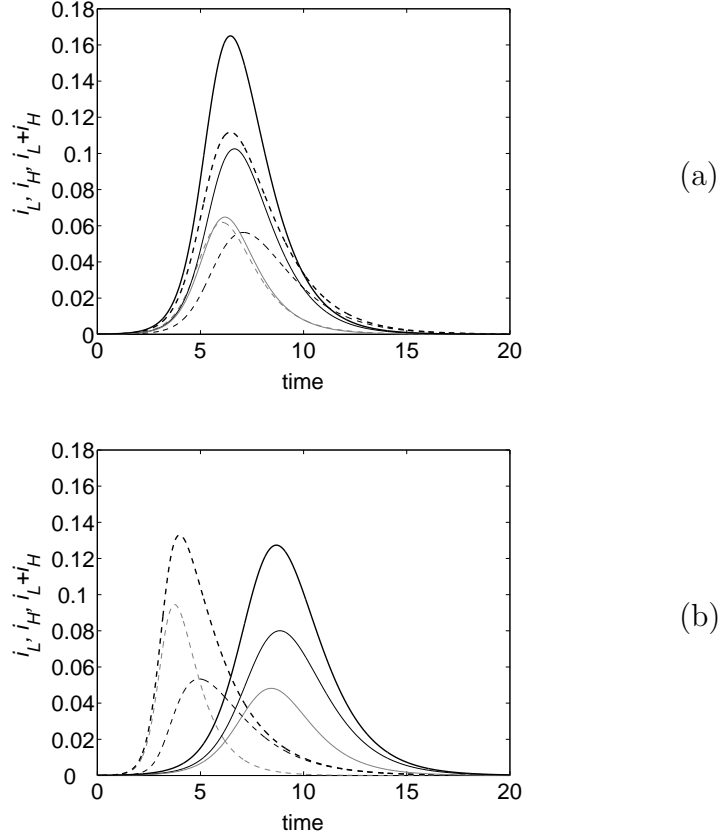


Figure 3: Proportion of the population infected as a function of time for disassortatively (solid lines) and assortatively (dashed lines) mixed populations. Cumulative proportion (thick lines,  $i_L + i_H$ ) and proportion of poorly (thin black lines,  $i_L$ ) and highly connected (thin grey lines, 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

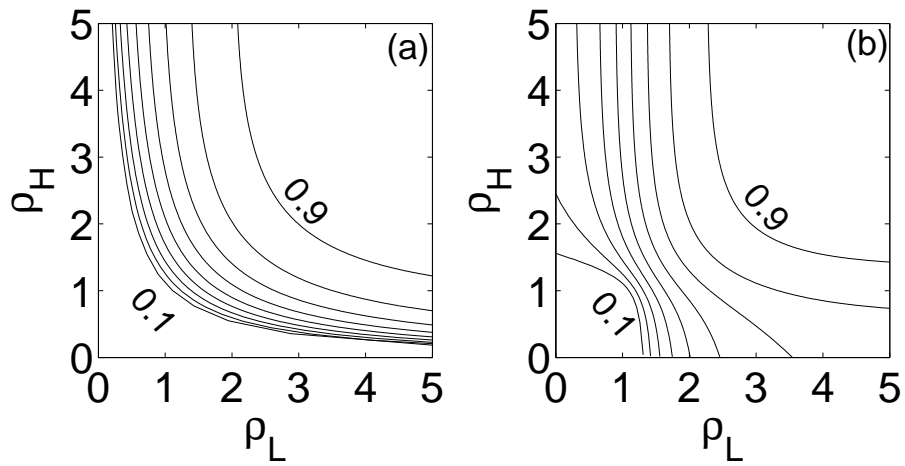
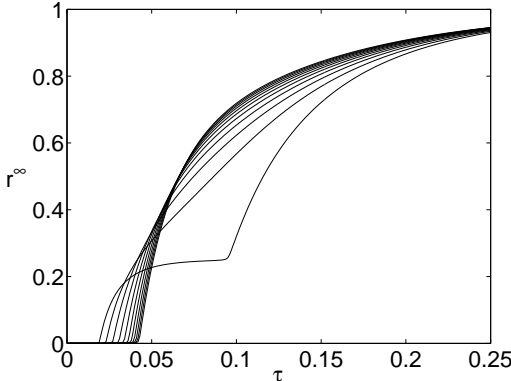
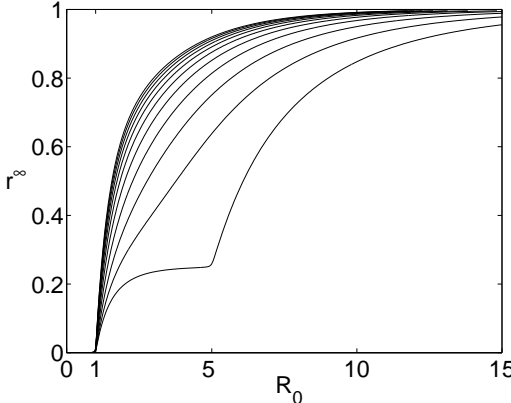


Figure 4: Final epidemic size as a function of the transmission potentials for (a) disassortatively-mixed ( $a = b = 0.05 \rightarrow r = -0.8$ ) and (b) assortatively-mixed ( $a = b = 0.45 \rightarrow r = 0.8$ ) populations. The case of  $f = 0.25$  is considered here.



(a)



(b)

Figure 5: Final epidemic size as a function of the per contact transmission rate  $\tau$  (a) and basic reproduction number  $R_0$  (b) for  $r = -1.0, -0.8, \dots, 1.0$  from left to right (a) and right to left (b). The case of  $a = b$  is considered here with  $k = 3$ ,  $\sigma = 5$ ,  $f = 1/(\sigma - 1) = 0.25$  and a value of  $\gamma \simeq 0.286$ .

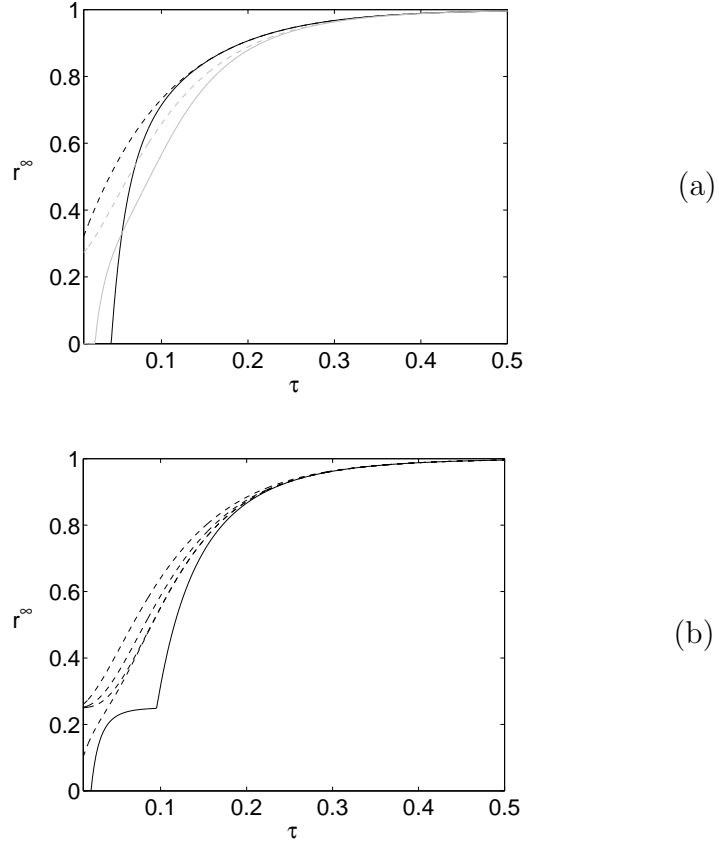


Figure 6: (a) Illustration of the approximating formula given by Eq. (41) (dashed lines) *versus* numerical solution (continuous lines) for  $r = -0.8$  (black lines) and  $r = 0.8$  (grey lines) (b) Illustration of the approximating formula given by Eq. (42) (dashed lines) with increasing number of terms *versus* numerical solution (continuous line). For both (a) and (b)  $k = 3$ ,  $\sigma = 5$ ,  $f = 1/(\sigma - 1) = 0.25$  and  $\gamma \simeq 0.286$ .