Mathematical Biosciences 231 (2011) 197–209

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Multiple sources and routes of information transmission: Implications for epidemic dynamics

Vasilis Hatzopoulos [a], Michael Taylor [a], Péter L. Simon [b], Istvan Z. Kiss [a,*]

[a] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton, BN1 9QH, UK
[b] Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary

ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form 8 February 2011
Accepted 1 March 2011
Available online 11 March 2011

Keywords:
Epidemic
Information transmission
Pairwise models
Simulation
Network overlap

ABSTRACT

In a recent paper [20], we proposed and analyzed a compartmental ODE-based model describing the dynamics of an infectious disease where the presence of the pathogen also triggers the diffusion of information about the disease. In this paper, we extend this previous work by presenting results based on pairwise and simulation models that are better suited for capturing the population contact structure at a local level. We use the pairwise model to examine the potential of different information generating mechanisms and routes of information transmission to stop disease spread or to minimize the impact of an epidemic. The individual-based simulation is used to better differentiate between the networks of disease and information transmission and to investigate the impact of different basic network topologies and network overlap on epidemic dynamics. The paper concludes with an individual-based semi-analytic calculation of R_0 at the non-trivial disease free equilibrium.

1. Introduction

As is evident all around us, human populations affected by the presence of an infectious disease will rarely remain passive. From leaflets found in a local clinic to community workshops and all the way up to national and global multimedia campaigns, concerted efforts are put into effect with the goal of changing peoples' behavior in the presence of an infectious pathogen [9,12]. The motivation for the above is that people, by virtue of the information content of these messages, will change their attitudes and actions to reduce their chances of becoming infected, spreading the disease further or experiencing prolonged periods of medical treatment. Often the messages in these campaigns are targeted to a particular geographic, sociodemographic (e.g. gender, age, ethnicity, income) or psychographic (e.g. those that are more likely to engage in risky behavior) sub-audience as many studies have found that the presence of a disease often correlates with an individuals' association with these segments [31]. When pathogens are known to spread through relatively well defined networks, as in the case of sexually transmitted diseases (STD), contact tracing data collected at clinics can be used to identify individuals with a prominent structural role in the network (e.g. hubs or bridges) who can then be approached or screened preferentially. Likewise individuals that are probable to influence opinion can be approached and seeded with information in the hope that they will diffuse this through their social network [24,31].

Besides such organized campaigns, observations suggest that people's decision on whether to adopt a change in behavior is based or influenced by others in their personal network of friends, colleagues and acquaintances [22,31]. This word-of-mouth effect has been observed and utilized in electronic marketing [13,23,27] and when modelling the diffusion of innovations [19], a paradigm that attempts to capture how people transition through the adoption process of a new product from non-aware to adopter (see [25] for a comprehensive review). Personal communications may also occur, although perhaps at a lower rate, between individuals that are not members of each others social network but come into contact rather infrequently and unpredictably. For example, an individual might unintentionally hear a conversation between two strangers on a bus. Such a 'mean-field' type of person to person transmission has been recently presented in [6] to model the transmission of infectious diseases between randomly and infrequently meeting individuals. Intuitively, such mean-field infections may be more frequent when considering an airborne pathogen than in STD's where the contacts are well determined and identifiable [15,29]. Besides media, social and random contacts a fourth route to awareness is an obvious one: being infected with a pathogen an individual will take measures to avoid infecting those in his/her neighborhood.

A major motivation for the incorporation of behavioral change models in infectious disease studies was the AIDS epidemic from the early 1980's and onwards. The main driver for this was the

* Corresponding author. Tel.: +44 1273873021; fax: +44 1273678097.
E-mail address: i.z.kiss@sussex.ac.uk (I.Z. Kiss).

0025-5564/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2011.03.006

realization that the growth of STD's, including AIDS/HIV, could be understood as a consequence of lifestyles choices and subjective risk perception motivated by individual attitudes, norms and beliefs [3,4].Recent behavioral work on HIV has concentrated on the growth of the epidemic in Africa which has been explosive [14,17,30]. In the past few years a number of compartmental epidemiological models have been proposed that incorporate various interactions of the diffusion of an infectious disease and human behavioral response. Broadly speaking most of the research can be classified into one of two categories. In the first class is models that deal with vaccine-preventable diseases (see [1,2,5,26] and references therein). In this case a natural question is whether (and to what extent) individuals' attitudes to vaccination can affect the dynamics of the disease. In cases of voluntarily vaccination, without further incentives, as vaccination coverage increases the risk of infection decreases due to herd immunity. At the same time any, real or imaginary, risks from the vaccine itself remain constant. This effect may motivate individuals to act in self-interest and avoid vaccination even if the risks of the vaccine are very small. Consequently disease eradication may become very difficult[26]. The second class of models deal with behavioral change in response to an epidemic outbreak [8,10,11,20,21,28]. Our model belongs in this class. Here individuals may alter the course of an epidemic by taking disease specific risk-reducing measures such as washing hands or using condoms. In modeling terms this is usually represented as subdivisions of a population into classes differentiated by degrees of risk exposure. It is beyond the scope of this paper to offer a review of this body of work but for the interested reader a comprehensive review can be found in [12]. Here the authors propose a classification based on the following criteria: the *source of information* - global or local, the *type of information* - prevalence or belief -based and finally the *effect of information*. The *type of information* is a classification meant to delineate whether behavioral change occurs due to the disease prevalence (prevalence-based) or due to the diffusion of some other behavioral trait that may be unrelated to the current prevalence (belief-based). The authors present as an example of this the decision of whether to vaccinate a child. Here, a conclusion might be reached without the disease in question being currently prevalent but based on a subjective perception of risks associated with vaccination. Finally the *effect of information* classifies how the presence of information alters an individuals exposure to the risk of infection. Possession of disease-related information might result in individuals changing their disease state via vaccination that can eliminate susceptibility, changing or influencing their contact network or taking measures to reduce the chances of acquiring or passing on infection. In light of this classification, we will present a model which encompasses both prevalence and belief-based types of information, local and global sources as well as the mean-field and self-induced avenues for information generation and transmission.

This paper builds on our previous work [20] where we proposed a compartmental model that coupled a simple *SITS* model with the diffusion of information generated by the presence of the disease. Here, we extend this model by introducing additional sources and routes of information transmission. We also provide a more fine-grained pairwise description of the problem along with an individual-based computational representation. The paper is organized as follows. In Section 2 the disease and information transmission models are introduced. These will serve as a basis for the pairwise model presented in the first part of Section 3 and the individual-based simulation model discussed in the second part. The pairwise model will be used to assess the potential of various routes of information generation and transmission to reduce the infectious prevalence as well as the benefits of using various combinations of these. In the second part of Section 3, simulations are used to increase the freedom of coupling or decoupling routes of disease and information transmission and to investigate the effect of network overlap for some simple network topologies. Finally, in Section 4, we present an individual-based analytic R_0 calculation at the non-trivial disease free steady state (DFSS) with further discussion in Section 5.

2. Model

Following on from [20], the population is divided into five different classes that specify the individual's status with respect to disease and information. These are: susceptible non-responsive (S_{nr}), susceptible responsive (S_r), infected non-responsive (I_{nr}), infected responsive (I_r) and in treatment (T). The term *responsiveness* emphasizes that the willingness to act or respond to the available information is key in trying to avoid infection or halting further spread. The important ingredients of the model relate to the *generation* and *transmission* of information as well as the *benefits of possessing and responding* to information. In the current model, information or responsiveness about the disease is generated in three ways: (**a**) $I_{nr} \to I_r$ as a result of symptoms, (**b**) $I_x \to T$, where $x \in \{nr,r\}$, as a result of being diagnosed and moving to the treatment class and (**c**) $X_{nr} \to X_r$, where $X \in \{S,I\}$ as a result of global information transmission. While the first two are intuitive, the latter is used to model the effect of mass-media campaigns which act as a single-source of information with its strength and duration often linked to the prevalence of infection in the population. Information transmission is possible in multiple ways and here we account for two: (**a**) locally or individual to individual and (**b**) mean-field. While disease dissemination locally represents the simple interaction of individuals where they can engage in discussions about an ongoing outbreak or diseases in general, the mean-field type transmission accounts for a less clear-cut interaction, such as overhearing a discussion, where information can be transmitted between individuals that are not necessarily in direct contact. Many of these different mechanisms of information generation and transmission can be easily linked to various ways in which information is disseminated in real-life.

It is natural to also incorporate the possibility for loss of responsiveness, consequently every X_r individual will transition to the X_{nr} class at a constant rate d_X, where $X \in \{S,I\}$. Note that this form is general enough that it could also model an individual's inability or refusal to act on information. Alternatively, as suggested in [20], the rate of information loss can be encoded as a decreasing function of the prevalence, a form that we do not explore in the current paper. The principal benefits of being informed and responding to the information can translate into reduced susceptibility, reduced infectivity and/or faster recovery if infected. For sake of completeness and to keep the model as general as possible, we incorporate all the above but we note that their presence of absence will depend on the precise modelling context and should be incorporated and used accordingly. The full suite of transitions are given in Table 1.

The present model, in some sense, can be seen as a generalization of the model in [10] but with a few important remarks. The model by Funk [10], is different in that information is only generated via self-diagnosis from infected individuals and this can lead to a qualitatively different behavior when compared to the present case where information is also generated via treatment. The model by Funk [10] uses a more general pairwise approach that accounts for two separate networks but it does not include the global prevalence-based transmission of information. In light of the above, the model proposed here shares some common features with some of the existing models but incorporate new ways

Table 1
All transitions allowed by the coupled infection/information system, where $X, Y \in \{S, I\}$ with individuals in treatment acting as members of the responsive classes (i.e. $X_r \in \{S_r, I_r, T\}$). Individuals in the treatment class return to being susceptible non-responsive and responsive at rate pr and $r(1-p)$ with $0 \leqslant p \leqslant 1$, respectively. The reduced susceptibility, infectivity and faster recovery, as a result of acting on information, is captured by the discount factors σ_s, $\sigma_i \in (0,1]$ and $\sigma_r > 1$. To model the mean-field transmission of information it is assumed that in unit time an individual may momentarily come into contact with k_{MF} others not in their social network. Along such links information flows at a rate m_X. The function $G_X([I_{nr}],[I_r])$ maps the prevalence of infection to the unit interval and is subsequently multiplied by the constant rate δ_X. This form models the saturating effect of media on individual behavioral response. Further explanations of the model rates and parameters are offered in the main text and in Appendix A.

Transition	Rate	Contact	Type
$I_{nr} + S_{nr} \to 2I_{nr}$	τ	G_d	Infection
$I_{nr} + S_r \to I_{nr} + I_r$	$\sigma_s\tau$	G_d	Infection
$I_r + S_{nr} \to I_r + I_{nr}$	$\sigma_i\tau$	G_d	Infection
$I_r + S_r \to 2I_r$	$\sigma_s\sigma_i\tau$	G_d	Infection
$I_{nr} \to T$	γ_{nr}	Independent	Infection
$I_r \to T$	$\sigma_r\gamma_{nr}$	Independent	Infection
$T \to S_{nr}$	$r{\cdot}p$	Independent	Infection
$T \to S_r$	$r{\cdot}(1-p)$	Independent	Infection
$X_r + Y_{nr} \to X_r + Y_r$	α_X	G_i	Information
$X_r + Y_{nr} \to X_r + Y_r$	$m_X k_{MF}$	Mean-field	Information
$X_{nr} \to X_r$	$\delta_X G_X([I_{nr}],[I_r])$	Independent	Information
$I_{nr} \to I_r$	ω	Independent	Information
$X_r \to X_{nr}$	d_X	Independent	Information

in which information can be generated and transmitted and gives further insight into the problem via semi-analytic R_0 calculations.

3. Results

We seek to explore the efficacy of our chosen mechanisms that model behavioral change in attempting to slow or stop the spread of a disease. This will be achieved by using a pairwise approximation (see Appendix A) and individual-based simulation model as well as a probabilistic semi-analytic R_0 calculation at the non-trivial DFSS. An understanding of the capabilities of these processes should provide useful information when designing information campaigns to fight a potential disease outbreak or when attempting to assess the impact of those that are currently implemented.

3.1. Pairwise model: comparison of information source efficacy

Pairwise ODE models [18] provide a good compromise between simple compartmental and full simulation models and allows us to capture more of the local nature of contacts and to depart from the very limiting homogeneous random mixing assumption. For example, this is important when modelling contact tracing [7,16] where control relies on being able to answer the 'who infected whom' type questions. In this case, the situation is similar in that the local nature of individual to individual transmission of information can lead to clusters of responsive individuals that are difficult to capture via simple compartmental models. Based on all possible transitions detailed in Section 2, pairwise equations can be derived heuristically or based on a mechanistic approach. In total there are 20 equations, 5 for singles and 15 for pairs (see Appendix A). The dimensionality of the system could be further reduced by taking into account that the population is closed and that all pairs add up to $\langle k\rangle N$, where $\langle k\rangle$ is the average node degree and N is the population size. The system of equations can be numerically integrated using standard methods (for more details and parameter values see Appendix A). The system exhibits three qualitatively different behaviors: (**a**) neither disease nor responsiveness can spread, (**b**) only responsiveness spreads and a state of endemic-responsiveness is reached and (**c**) both responsiveness and infection are endemic.

The analysis begins by comparing information sources with respect to their capacity to bring the prevalence to a desired low level, when each is acting in isolation. Information can achieve this by shifting a large fraction of the population into the responsive class. Such informed individuals will then experience decreases in their infectivity and susceptibility as well as a faster recovery. The minimum rate at which a source of information can bring the prevalence to a desired low level will be referred to as the *critical information transmission rate*. We choose a prevalence level greater than zero, specifically $I^{eq} = 0.01$, because as the prevalence approaches zero precise identification of the critical information rates becomes numerically challenging. Starting with a small initial information generation or transmission rate, for a range of infection rates τ with fixed recovery parameters, the system is seeded with a small number of I_{nr} and S_r individuals. The system of pair approximation equations (see Appendix A) is then numerically integrated to identify the smallest or critical rate that will lead to the desired prevalence level I^{eq}. Next, the relative capacities of α, ω and δ to deliver a state of low infection prevalence for different values of τ are considered. We let $p = 0.9$, which approaches a worse case scenario limit whereby no information is generated by the individuals themselves through past experience. This parameter choice allows us to examine the effects of α, ω and δ in relative isolation. As shown in Fig. 1 contact-based transmission of information is by far the most efficient way to generate a responsive population, a result well known in the diffusion of innovations literature [25]. In this case every receiver of information (I_{nr} or S_{nr}) immediately also becomes a transmitter of information, in contrast to global transmission of information where the source of information is at all times singular. The mean-field type transmission of information, not shown in Fig. 1, is equally potent and produces results that are similar to the contact-based transmission case, especially if the network is densely connected. For smaller values of $\langle k\rangle$, the mean-field transmission performs better than the purely contact-based but the differences are small.

We model the transition to the responsive class due to media exposure at a rate given by the function

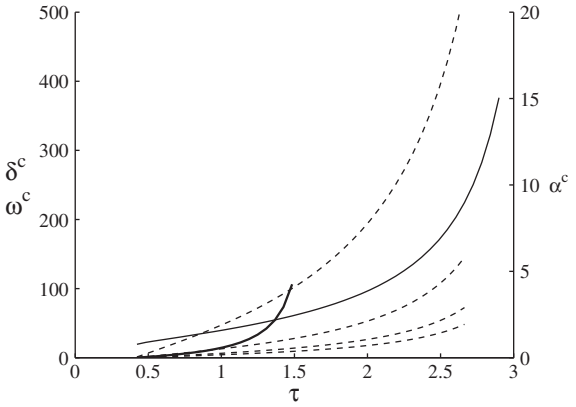

Fig. 1. Critical information rates resulting in a prevalence of $I^{eq} = 0.01$ as a function of the per-contact infection transmission rate τ (computed via the pairwise model). For $\tau > 0.42$ and in the absence of information, the prevalence equilibrates at $I^{eq} > 0.01$. At and beyond this point different amounts of each information rate are needed to lower the prevalence to $I^{eq} = 0.01$. In this case, the effect of each transmission route is investigated in the absence of all others. Solid and thick solid line correspond to α^c and ω^c, respectively. The four dashed lines represent δ^c for different values of the population inertia parameter $K = [0,5,20,100]$ increasing from right towards the left. The values for α^c are denoted on the right y axis and for all other rates on the left y axis. Other parameters fixed across all data points are $p = 0.9$, $\sigma_s = \sigma_i = 0.5$, $\sigma_r = 2$, $\gamma_{nr} = 2$, $\gamma_r = \sigma_r\gamma_{nr}$, $d = \gamma_r$, $k = 6$.

$$G_s([I_{nr}],[I_r]) = G_i([I_{nr}],[I_r]) = \frac{\delta([I_{nr}]+[I_r])^n}{K+([I_{nr}]+[I_r])^n}, \quad (1)$$

where in this paper, $n = 1$ at all times. The effectiveness of global information (acting on I_{nr} or S_{nr}) is strongly tied to the K parameter which controls the growth of $G()$ such that when the prevalence is low the function grows like $\frac{1}{K}(I_{nr}+I_r)^n$. It is helpful to think of K as a measure of population inertia in responding to information (note that it is also possible to include such inertia parameters in the network based information rates). Populations with high values of K are resistant to behavioral change which can therefore act as an indicator for the quality of global information campaigns. For example, high values of K will simply lead to observing vanishingly small benefits from global information campaigns. The critical rates for self-diagnosis are at best similar to those for global information, especially for diseases with low transmissibility. As is the case for global information, self-diagnosis will only function once infected individuals are present and the situation is somehow even less fortunate given that ω can only act on I_{nr}. The self-diagnosis rate can be thought of as a model for the probability of a particular infection being symptomatic. Thus for diseases with mild symptoms or those that are asymptomatic the need for peer-to-peer communication and low population inertia is even more pressing. Finally self-diagnosis can, in cases where the population has very high behaviorial inertia, be more effective than global information dissemination. This can be seen by comparing the appropriate curves in Fig. 1. It is also worth noting that as τ increases it is less and less likely that information generation and/or transmission can prevent an epidemic. More precisely at large but finite values of τ, the rates of information generation and/or transmission needed to halt the spread will tend to unfeasible large values.

The numerical and individual based simulations suggest that global information is never a more efficient way to lower infection prevalence than contact-based transmission. The global information transmission rate takes its maximum value as $K \to 0$, for large δ and high prevalence of infection, at which point $G \simeq \delta$. For any value of δ we have found that there is a contact-based rate $\alpha(\alpha < \delta)$ that will lower the prevalence at least by the same amount. Preliminary results also suggest that increasing the node/individual specific heterogeneity in σ, while keeping the same mean, results in the lowering of prevalence as a large number of nodes, with small values of σ, are almost immune or unable to transmit the infection. A somewhat similar observation holds for α, where many nodes with limited potential to transmit the information will lead to a higher prevalence.

In reality no single information source will act in isolation. Media campaigns encourage discussion which can bring forth behavioral change. Infected individuals are likely to learn from experience and further communicate this to their family and friends. Our model is able to accommodate such scenarios as we show in Fig. 2 for various combinations of α, δ and ω. As expected the α and δ combination is the most effective pair while the combination of all three sources is capable of eradicating a large majority of epidemics for large τ. Contact based transmission of information is by far the most efficient as it constantly builds new sources of information. Responsive individuals are then able to halt the spread of an epidemic by forming clusters that can resist infection invasions and this is discussed in more detail in the next section.

3.2. The impact of network overlap on prevalence equilibria

In the most general case and for a given host population, infection diffuses on a network G_d and the social interactions can be represented by a network G_i. The information transmission dynamics coupled with the overlap pattern between networks determine

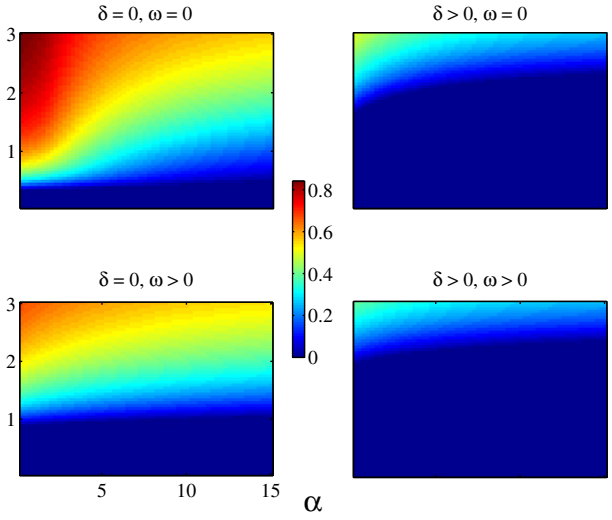

Fig. 2. The effect of combining different sources of information. On the top left panel the endemic infection prevalence is shown for a range of α and τ values. In the remaining panels, for each combination of α and τ, either global information or self-diagnosis or both are added with the same constant rate equal to 12. All other parameters are the same as in Fig. 1. (For interpretation to colours in this figure, the reader is referred to the web version of this article.)

the extent to which disease can be slowed or halted. Slowing or stopping an epidemic depends on both the number of responsive individuals as well as their precise distribution around sources of infection. Hence, increasing the effectiveness of the information/responsiveness dissemination is equivalent to an optimization process where increasing the number of responsive individuals has to be coupled with achieving an optimal correlation pattern. For example, it is beneficial if responsive susceptibles cluster around infectives. Here, we focus on investigating how network overlap and different information generation mechanisms impact on these correlation patterns and ultimately on model outcome. The role of the spatial structure or correlations is explored via individual-based stochastic simulation. As in the case of the pairwise model, the simulation model admits three equilibria: the trivial disease-free steady state (DFSS), the non-trivial DFSS and a state where information and infection remain endemic. Here, we aim to examine the mixing patterns between S_{nr}, S_r, I_{nr} and I_r in the endemic state and between S_{nr} and S_r in the non-trivial DFSS. Furthermore, these correlations will be compared between non-overlapping and overlapping networks.

The growth of infection is shown in Fig. 3 for Poisson random graphs and for different methods of information generation (i.e. via treatment alone with $p = 0$ and $\omega = 0$ and via self-diagnosis with $p = 1.0$ and $\omega \neq 0$). In both cases, once information is generated it can only be passed on via the information transmitting network. The prevalence of infection grows faster when networks do not overlap, although this difference is more significant in the $\omega \neq 0$ case. This is in line with previous findings [10] but merits closer scrutiny. In the $\omega \neq 0$ case, information is generated via $I_{nr} \to I_r$ type transitions and thus the newly informed infectives can transmit both information and disease. Given that the neighbouring nodes on G_d of these I_r individuals are at immediate risk of infection, it is desirable that these nodes are also neigbours of the I_r individuals on G_i. If this is the case, and the transmission rate of information is faster than the transmission rate of the disease, it is likely that infection can be stopped. However, if the networks do not overlap, then responsive or informed individuals will transmit the information to those that are not at immediate risk of infection and thus with little immediate benefit. Therefore, in this

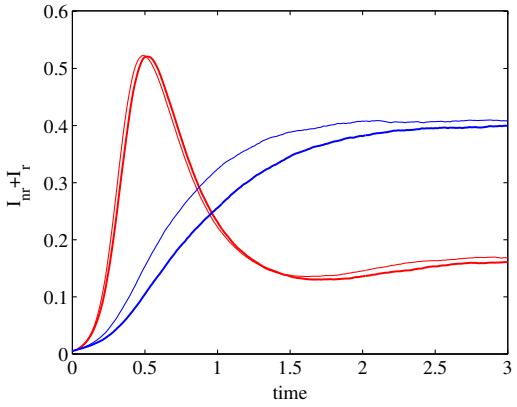

Fig. 3. Time evolution of the total infection prevalence, based on simulation results, for overlapping (thin lines) and non-overlapping (thick lines) Poisson random graphs. Responsiveness is generated by those in treatment with $p = \omega = 0$ (red lines) and by infected individuals with $p = 1$ and $\omega = 52$ (blue lines). In all cases, responsiveness only diffuses via neighborhood contacts. When the networks do not overlap the number of infected grows faster, however when responsiveness is generated by individuals in treatment the overlap plays a less significant role. Both systems tend to very similar states with infection prevalence being slightly higher in the non-overlapping cases. Other parameters are: $\tau = 2.5$, $\alpha = 1.3$, $\sigma_s = \sigma_i = 0.5$, $\sigma_r = 2, \gamma_{nr} = 2$, $\gamma_r = \sigma_r\gamma_{nr}$, $d = \gamma_r$, $N = 10^4$ and $\langle k\rangle = 8$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

case, information acts similarly to random vaccination with benefits only coming into play when infection reaches pockets of responsive clusters. According to Fig. 3, complete overlap slows the epidemic initially but it cannot preemptively inform individuals that are not at immediate risk of infection. For complete non-overlap however, the situation changes in that the epidemic picks up early but later on reaches a large number of individuals with reduced susceptibility and therefore the spread slows. While the long term behaviour for non-overlap and overlap is similar, the slow initial growth may offer an advantage as it extends the period to implement other complementary control measures.

On the other end of the spectrum, information could be generated by those in treatment. In this case the amount of edge overlap between the network has a less significant effect on the time evolution of the spread and final outcome. Here, individuals in treatment (Ts) can typically inform S_{nr}s and/or I_{nr}s. However, even for complete overlap, informing non-responsive susceptible neigbours is not of immediate benefit since these are not at immediate risk of infection. The most significant benefit of complete overlap is the transmission of information from Ts to I_{nr}s. This will lower the infectivity of infectious individuals but it is far less effective compared to the $\omega \neq 0$ case due to information lagging behind the transmission of the disease. In effect, this is similar to contact tracing where, new infections are identified via the infector. In this case, the slight differences between overlap and non-overlap (see Fig. 3) is also reflected in the time evolution of the $[TI_{nr}]$ pairs whereby the growth of these is faster in the non-overlapping case. This means that more I_{nr}s remain uninformed and thus lead to a slightly faster initial increase in the prevalence of infection. This effect disappears quickly as the information makes progress and many individuals become responsive. At this stage the local interaction has less of an effect given that a considerable part of the individuals are already responsive.

In the context of our model, the arguments above can be underpinned by investigating the propensity of susceptible responsive individuals to cluster around infectives. This can be measured by the following conditional probability

$$P(S_r - I_y|S_x - I_y) = \sum_{y=r,nr}[S_rI_y]/\sum_{x,y=r,nr}[S_xI_y].$$

This represents the fraction of all $S_x - I_y$ pairs with $S_x = S_r$ and indicates the correlation of information and infection. If information and infection are neutrally correlated, for example if they spread on two networks that have disjoint edge sets, then one would expect $P(S_r - I_y|S_x - I_y) = Q(S_r) = N_r/N$ (i.e. equal to the probability of finding an S_r if picking nodes at random). The time evolution of the conditional probabilities for overlapping and non-overlapping Poisson networks are shown in Fig. 4. The correlations develop and persist as the systems settle to an equilibrium prevalence. The amount of overlap plays a significant role only when information is generated via self-diagnosis (i.e. $\omega \neq 0$) (see blue lines in Fig. 4), and in this case, information is close to infection as shown by higher values of the correlations. On the other hand if the infectives spread infection and responsiveness to a completely disjointed neighbourhoods then initially there is very little correlation until later when there is a substantial amount of infection and information. In all cases the system settles to an endemic state with non-zero infection and responsiveness and will remain correlated (i.e. conditional probability far from the diagonal). In this equilibria, there is high correlation between infectives and informed since both managed to invade and occupy parts of the network. This particular correlation is a result of the system being in a form of statistical equilibrium where information loss and gain as well as recovery and new infections all balance out. Fig. 4 also shows the good qualitative agreement between the pairwise and simulation model for complete overlap, and highlights the merits of concurrently using both.

Using pair-wise models permit us to track the evolution of pairs that capture the spatio-temporal diffusion of information at a level of description unattainable in simple ODE models. Furthermore pair-approximation are naturally more appropriate descriptions of simulation-based network models which are based on explicit structure representations[16]. As will shall see in the next section knowledge of the information mixing patterns will prove instrumental in performing a novel individual based R_0 calculation.

4. Individual-based calculation of R_0^d

The coupled disease and information transmission model admits two disease-free steady states: (a) trivial (1,0,0,0,0) and (b)

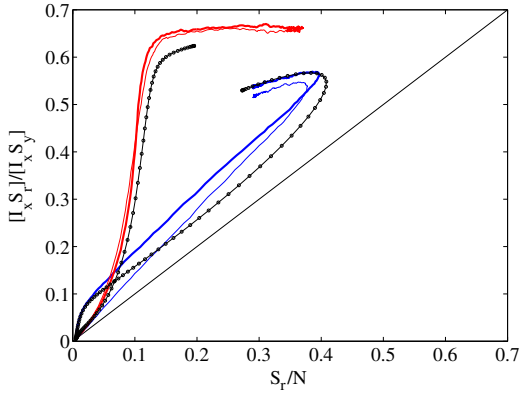

Fig. 4. Awareness in susceptible members of $S - I$ pairs as a function of the total awareness prevalence in the population, S_r/N. Results based on simulations on Poisson random graphs for full overlap (thick lines) and no overlap (thin lines). Responsiveness is generated by those in treatment with $p = \omega = 0$ (red lines) and by infected individuals with $p = 1$ and $\omega = 52$ (blue lines). Data from the pairwise model is shown by the circled lines. This correlation measure distinguishes between the two overlap limits only when information generation is by the infected. In both cases correlations remain above the diagonal indicating that the system reached an equilibrium where neither infection nor responsiveness spread further. All other parameters are the same as in Fig. 3.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

non-trivial $(1-s_0, s_0, 0, 0, 0)$ (DFSSs). The trivial DFSS can be perturbed via the spread of infection and/or responsiveness, provided that system is seeded accordingly or whether responsiveness can be generated directly (i.e. $I_{nr} \to I_r$). Here, the case of the trivial disease-free steady state is not discussed (see [11]) and the focus is on determining the potential of an initial infected individual to invade or spread within a population that is at an equilibrium with respect to the transmission of responsiveness. Thus, the basis for calculating R_0^d is provided by the characteristics of the equilibrium of endemic responsiveness, where all nodes are either S_{nr} or S_r and are arranged in a particular configuration on the disease transmission network (G_d) as determined by the information transmission process (e.g. local via contacts G_i and mean-field) and the particular overlap between the routes of responsiveness and disease transmission.

In this case, computing R_0^d amounts to determining whether a particular equilibrium is stable against perturbation induced by mutating a randomly chosen individual to the I_x state, where $x \in \{nr, r\}$. A population with elevated levels of responsiveness will be able to slow or halt the spread of an epidemic by virtue of reductions in the infectivity, susceptibility and an increase in the recovery rate of responsive individuals. Consequently, in this model we define R_0^d as: *the expected number of secondary infections caused by a typical infective in a completely susceptible population in a state of endemic responsiveness*. In general, R_0^d will depend on the amount of responsiveness but also on the presence, or absence, of correlations between the non-responsive and responsive individuals/nodes. The R_0^d calculations assume that the distribution responsiveness is at equilibrium, hence neighbors of the initial index case will not change their responsiveness status due to forgetting or become responsive due to responsive neighbors on the information transmission network G_i.

The crucial ingredients for determining R_0^d are: (a) the frequency of S_{nr} and S_r individuals denoted by $Q(S_x) = [S_x]/N$, where $x \in \{nr, r\}$ and (b) the correlation pattern, and in particular, the conditional probabilities (i.e. $P(S_y|S_x)$ with $x \in \{nr, r\}$) that describe the neighborhood composition of a node on the G_i network as given by the population level average, at the state of endemic responsiveness. The first step in calculating R_0^d is to select at random a susceptible individual and mutate this into an I_{nr} or I_r individual. Thereafter, the spread of the disease will depend on the local neighborhood around the mutated individual and the enumeration of all the possible ways in which new infections can be generated. Information about the average number of possible onwards infections is provided the average excess degree of a node, $D = \langle k \rangle - 1 + \frac{\mathrm{Var}(k)}{\langle k \rangle}$, where $\langle k \rangle$ is the average number of connections per individual and Var (k) represents the variance as given by the node degree distribution on the G_i network. Combining all the above, the expression for R_0^d in the case of proportionately or randomly mixed networks can be written as

$$R_0^d = D \sum_{x,y,z \in \{nr,r\}} Q(S_x) P(S_x \to I_y) P(S_z|S_x) P((I_y, S_z)), \tag{2}$$

where $Q(S_x)$ represents the probability of picking a node of type S_x at random and $P(S_x \to I_y)$ represents the probability of mutating a susceptible node to I_x, $x \in \{nr, r\}$. This probability is arbitrary and is not an endogenous system parameter (i.e. one could chose any arbitrary value in $[0,1]$). $P((I_y, S_z))$ denotes the probability with which an (I_y, S_z) edge generates a new infection. Expanding the sum we get the following eight terms:

$$\begin{aligned}
R_0^d = {} & Q(S_{nr}) P(S_{nr} \to I_{nr}) P(S_{nr}|S_{nr}) P((I_{nr}, S_{nr})) \\
& + Q(S_{nr}) P(S_{nr} \to I_{nr}) P(S_r|S_{nr}) P((I_{nr}, S_r)) \\
& + Q(S_{nr}) P(S_{nr} \to I_r) P(S_{nr}|S_{nr}) P((I_r, S_{nr})) \\
& + Q(S_{nr}) P(S_{nr} \to I_r) P(S_r|S_{nr}) P((I_r, S_r)) \\
& + Q(S_r) P(S_r \to I_{nr}) P(S_{nr}|S_r) P((I_{nr}, S_{nr})) \\
& + Q(S_r) P(S_r \to I_{nr}) P(S_r|S_r) P((I_{nr}, S_r)) \\
& + Q(S_r) P(S_r \to I_r) P(S_{nr}|S_r) P((I_r, S_{nr})) \\
& + Q(S_r) P(S_r \to I_r) P(S_r|S_r) P((I_r, S_r)).
\end{aligned} \tag{3}$$

Assuming that a state of endemic responsiveness has been established in the population, $Q(S_x)$ and $P(S_y|S_x)$ are given. We note that the expressions for the conditional probabilities $P(S_y|S_x)$ are independent of node degree. However correlations between the information compartments and degrees of two nodes at the ends of a link can be introduced by making the generalization $P(S_y|S_x) \to P(S_y^l|S_x^k)$ where l,k are node degrees that are bounded by the network connectivity. Strictly speaking the calculations presented here work best when the network is homogeneously mixed. The probability of mutating into I_x can be binary or imposed as required by a particular modelling context. The more technical part is to write down the probability that infection will happen across an (S_x, I_y) edge. All these calculations rely on a schematic diagram as shown in Fig. 6. Here, starting with a particular pair (i.e. (S_{nr}, I_{nr})), all potential end states are listed by following a Markovian type process with one single change allowed during any single step. The complications arise from the fact that self-diagnosis or the repeated forgetting and acquiring of responsiveness leads to an infinite tree of possible outcomes which makes calculations non-trivial. A new infection on any such tree results will result from an (S_{nr}, I_{nr}), (S_{nr}, I_r), (S_r, I_{nr}) or (S_r, I_r) pair denoted by *A*, *B*, *C* and *D*, respectively. To fully describe the problem, all trees starting with any one of the four pairs needs to be considered. Independently of the source of the tree, all four pairs will be present in future generations, and calculating the probabilities of encountering these pairs is crucial for calculating R_0^d. Before discussing the four different tree types, it is worth noting that the probabilities of an infection originating from pairs *A*, *B*, *C* and *D*, given all other possible events, can be simply written as

$$\begin{aligned}
& p_A^{inf} = \frac{\tau}{\gamma_{nr} + \omega + \tau}, p_B^{inf} = \frac{\tau \sigma_i}{\gamma_r + \alpha + \tau \sigma_i + d}, \\
& p_C^{inf} = \frac{\tau \sigma_s}{\gamma_{nr} + \omega + \alpha + \tau \sigma_s + d}, p_D^{inf} = \frac{\tau \sigma_s \sigma_i}{\gamma_r + 2d + \tau \sigma_s \sigma_i}.
\end{aligned} \tag{4}$$

For a given starting point in the tree (say *A*), the first task is to obtain the *expected number of pair types* $A\left(P_A^A\right), B\left(P_B^A\right), C\left(P_C^A\right)$ and $D\left(P_D^A\right)$ that can be reached in the complete tree. These expectation values can then be simply multiplied by the probability of a new infection resulting from a particular pair (i.e. $\sum_{X \in \{A,B,C,D\}} p_X^{inf} P_X^A$). Below, for the tree starting with node *A*, the full details of the calculations are given followed by more succinct summaries for the three remaining trees in Appendix D. The technique only depends on the precise knowledge of all possible transitions from a given pair to other pairs and the probability of these transitions as listed in Appendix B.

4.1. Tree starting with pair (S_{nr}, I_{nr})

The removal of all end states from Fig. 5 results in a simplified diagram (see Fig. 6(a)). To work out the probability of reaching all *A*, *B*, *C* and *D* pairs, the tree is first simplified by noting that in every even generation, with the starting *A* being generation 0, only *A*s and *D*s occur, and similarly *B*s and *C*s only appear in every odd generation. Hence, the tree can be further reduced to a tree with only *A*s and *D*s as given in Fig. 6(b). It is now useful to define the probability of reaching a particular pair type i in generation n, given that the tree started with pair X, by $i^X(n)$, where $i \in \{a, b, c, d\}$, $X \in \{A, B, C, D\}$ and $n = 0, 1, 2, \ldots$. Based on this, and on the

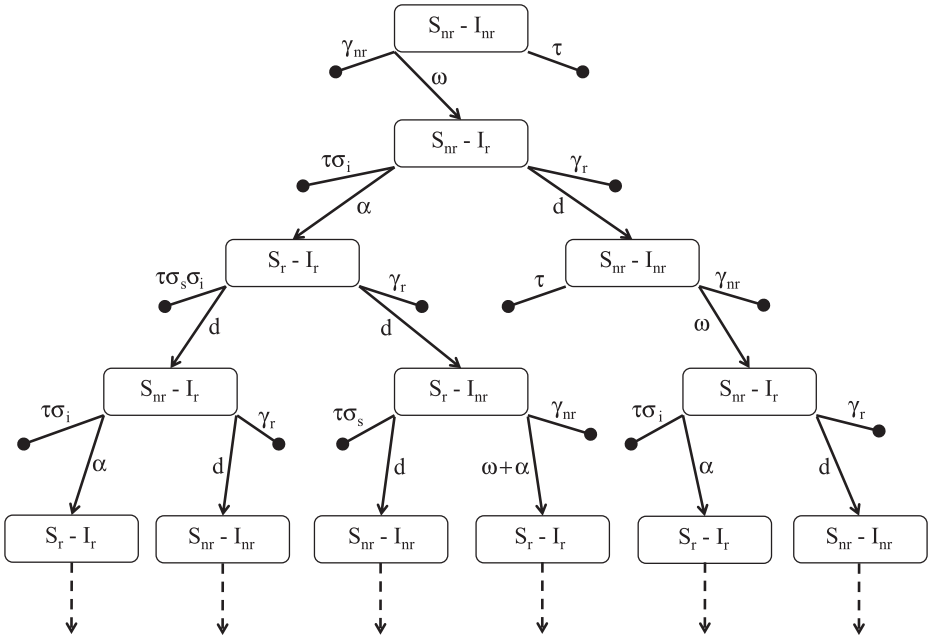

Fig. 5. The diagram/tree of all transitions starting from pair (S_{nr}, I_{nr}). This is constructed based on a Markov process type transition with only one change allowed at each step.

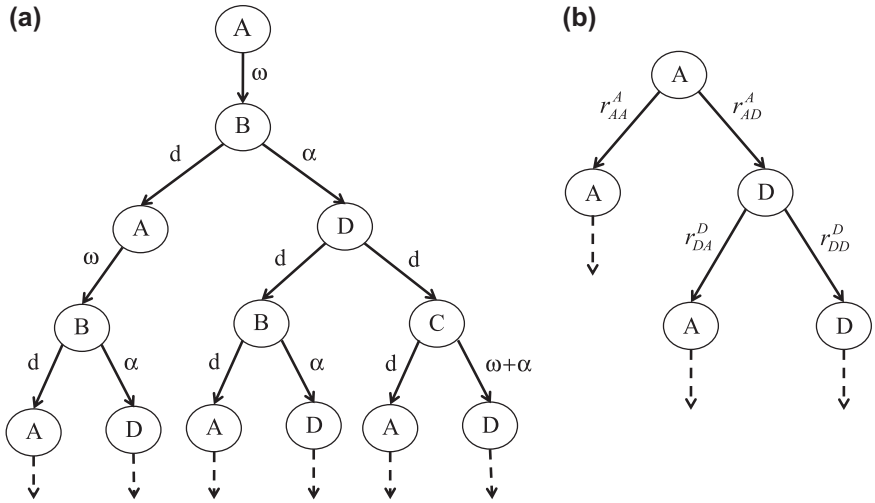

Fig. 6. (a) The equivalent of Fig. 5 based on excluding all end states and denoting pairs (S_{nr}, I_{nr}), (S_{nr}, I_r), (S_r, I_{nr}) and (S_r, I_r) by A, B, C and D, respectively. (b) A reduction of the tree in (a) obtained by only considering pairs in every even generation, with the tree starting at generation zero. The rates of the reduced tree, r^A_{AA}, r^A_{AD}, r^A_{DA} and r^A_{DD}, are obtained from the original tree by enumerating all possible ways of getting the desired transitions together with their probabilities.

simplified tree in Fig. 6(b), the following system of coupled recurrence equations can be derived,

$$a^A(n+1) = r^A_{AA} a^A(n) + r^A_{DA} d^A(n), \tag{5}$$

$$d^A(n+1) = r^A_{AD} a^A(n) + r^A_{DD} d^A(n), \tag{6}$$

with initial condition $(a^A(0), d^A(0)) = (1,0)$, where $r^A_{AA} = r_{AB} r_{BA}, r^A_{AD} = r_{AB} r_{BD}, r^A_{DA} = r_{DB} r_{BA} + r_{DC} r_{CA}$ and $r^A_{DD} = r_{DB} r_{BD} + r_{DC} r_{CD}$. We note that these probabilities are derived based on examining the full tree and considering the probability with which As and Ds generate more As and Ds two generations down the tree. Working out the probability of going from a node to another simply amounts to multiplying together all transition probabilities along the path from a source to a target node. For the first few generations, it is easy to check that $a^A(n)$ will give the sum of all probabilities of reaching all possible As present in generation n. Hence, the following identities hold

$$P^A_A = \sum_{n=0}^{\infty} a^A(n), \quad P^A_D = \sum_{n=0}^{\infty} d^A(n). \tag{7}$$

Given that each A generates a B with probability r_{AB}, and that D generates a B and a C with probability r_{DB} and r_{DC}, the following equations hold

$$P^A_B = P^A_A r_{AB} + P^A_D r_{DB}, P^A_C = P^A_D r_{DC}. \tag{8}$$

Therefore, the only task is to find P^A_A and P^A_D. Using simple linear algebra techniques, as detailed in Appendix C (see Eqs. (47) and (48)), these are given by

$$P^A_A = \frac{1 - r^A_{DD}}{1 - (r^A_{AA} + r^A_{DD}) + r^A_{AA} r^A_{DD} - r^A_{AD} r^A_{DA}},$$

$$P^A_D = \frac{r^A_{AD}}{1 - (r^A_{AA} + r^A_{DD}) + r^A_{AA} r^A_{DD} - r^A_{AD} r^A_{DA}}. \tag{9}$$

Collating all the information above, the probability of an infection from an (S_{nr}, I_{nr}) pair is

$$P((S_{nr}, I_{nr})) = \sum_{X\in\{A,B,C,D\}} p_X^{inf} P_X^A. \tag{10}$$

The derivation of the trees starting with states B, C and D in generation 0 are very similar and are shown in Appendix D. All the results from the four starting pairs feed directly into Eq. (2) to provide an analytical expression for R_0^d. This expression provides the means to investigate and tease apart the impact of factors such as the number of responsive individuals as well as the precise correlation between non-responsive and responsive individuals. The latter can be consolidated in a correlation matrix given as,

$$P_{corr} = \begin{pmatrix} P(S_{nr}|S_{nr}) & P(S_r|S_{nr}) \\ P(S_{nr}|S_r) & P(S_r|S_r) \end{pmatrix}.$$

Working under the assumption of homogeneous random mixing, distributing $Q(S_{nr})N$ non-responsive and $Q(S_r) = N(1 - Q(S_{nr}))$ responsive individuals results in a correlation pattern given by $P(S_{nr}|S_{nr}) = P(S_{nr}|S_r) = Q(S_{nr})$ and $P(S_r|S_{nr}) = P(S_r|S_r) = Q(S_r)$. In Fig. 7(a), the value of R_0^d is plotted for different proportions of non-responsive individuals. While in this case both the number of responsive individuals and correlation change, the effect on R_0^d is significant with a higher number of responsive individuals leading to a significantly lower value of R_0^d. In Fig. 7(b), the number of non-responsive individuals is kept fixed and only the correlation pattern is varied. This can be achieved using a simple two parameter model that determines uniquely the state of endemic responsiveness. In particular, for a regular random graph, considering $x = Q(S_r)$ as the first free parameter the following equations must hold

$$Q(S_{nr}) = 1 - x, \tag{11}$$

$$P(S_{nr}|S_{nr}) + P(S_r|S_{nr}) = 1, \tag{12}$$

$$P(S_{nr}|S_r) + P(S_r|S_r) = 1, \tag{13}$$

$$Q(S_{nr})P(S_r|S_{nr}) = Q(S_r)P(S_{nr}|S_r), \tag{14}$$

where, the first equation is trivial, the second and third are the normal requirements for conditional probabilities and the fourth is the detailed balance condition which guarantees that individuals of different types are connected up in a consistent manner. The second free parameter is chosen to be $y = P(S_{nr}|S_{nr})$, and all other variables follow in terms of x and y,

$$P(S_r|S_{nr}) = 1 - y, P(S_{nr}|S_r) = \frac{(1-x)(1-y)}{x} \text{ and}$$

$$P(S_r|S_r) = 1 - \frac{(1-x)(1-y)}{x}.$$

It is worth noting that all variables must be greater or equal zero and less or equal one. This requires that $x \in [0,1]$ and max $(0, (1-2x)/(1-x)) \leqslant y \leqslant 1$. Fig. 7(b) shows that for a fixed level of responsiveness in the population (see group of black and red curves), different correlations lead to significantly different values of R_0^d. It is not straightforward to establish a link between the correlation matrix P_{corr} and the value of R_0. However, some simple observations hold. For example if $P(S_{nr}|S_{nr})$ is high then the values of R_0^d will be high and, in certain regimes, increasing assortativity in responsiveness (e.g. non-responsive and responsive individuals preferentially connect to individuals of the same type) also leads to higher values of R_0^d. The lack of a simple relation is to be expected given that the impact of a particular pair type strongly depends on other parameters such as transmission or loss of responsiveness and the benefits of being in a responsive state.

Even though a recipe for the optimal spatial correlation of S_{nr}s and S_rs is not forthcoming, it is possible nonetheless to gain some insight into the effectiveness of certain configurations by examining the output from individual based simulations. In Fig. 8, the time evolution of the neighborhood around an informed susceptible for completely overlapping and non-overlapping Poisson random graphs is shown. In this case, information halts the spread of infection and the system settles to an endemic responsiveness. Fig. 8, for full network overlap, shows that S_r individuals quickly build an assortative mixing pattern which stays fairly constant over time and then quickly decays just above neutral mixing after the infection has died out. For non-overlapping networks the assortativity in S_r slowly builds up to a lower magnitude before it collapses on the diagonal. Although this behavior might seem to favor the overlapping networks, the less than expected dissasortativity in the overlapping case can hamper the spread of responsiveness (left panel Fig. 8). This is because it is now less likely to find a non-responsive next to a responsive type, a configuration which does not benefit the diffusion of information. In other words in completely overlapping networks information might get stuck in self sustaining clusters that can resist invasion whilst non-overlapping networks lead to almost neutral mixing in information which can aid diffusion at remote parts in the network, with both scenarios having their own merit.

The analytic expression of R_0^d can also be used to determine numerically the critical values of parameters such as α, ω and δ needed to bring the basic reproduction number below one. This might prove to be a useful tool in real applications where the level

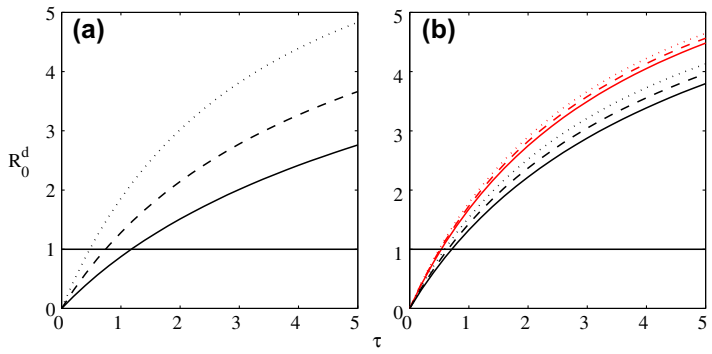

Fig. 7. Plots of R_0^d for (a) homogeneous random mixing with $Q(S_r) = 0.9$ (continuous line), $Q(S_r) = 0.5$ (dashed line) and $Q(S_r) = 0.1$ (dotted line), and for (b) fixed frequency of $Q(S_r) = 0.20$ and correlations given by $y = 0.75, 0.875, 1$ (black continuous, dashed and dotted lines), and frequency $Q(S_r) = 0.40$ with correlation patterns given by $y = 1/3, 2/3, 1$ (red continuous, dashed and dotted lines). The other parameter values are: $P(S_{nr} \to I_{nr}) = P(S_r \to I_r) = 1$, $P(S_{nr} \to I_r) = P(S_r \to I_{nr}) = 0$, $\sigma_s = \sigma_i = 0.50$, $\gamma_{nr} = 2.0$, $\gamma_r = 4.0$, $\omega = 2.0$, $d = \gamma_{nr}$ and $\alpha = 2d$.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

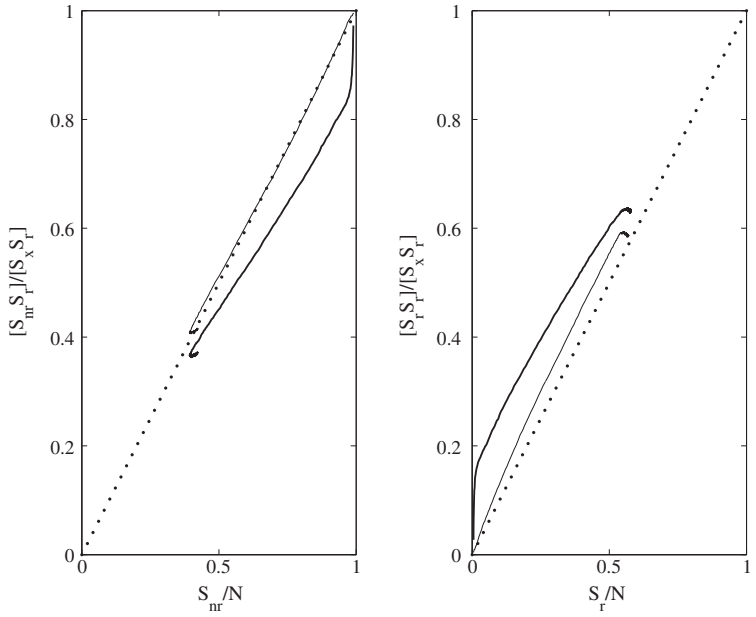

Fig. 8. The probability that in an $S_x - S_r$ pair we find $S_x = S_{nr}$ (left panel) and $S_x = S_r$ (right panel) for overlapping (thick lines) and non-overlapping (thin lines) Poisson random graphs plotted against the prevalence of the type S_x as the system evolves to its steady state. Information generated by those in treatment with $p = 0.5$ and $\omega = 0$. All other parameters are the same as in Fig. 3 except $\tau = 0.8$.

and impact of control interventions need to be measured. The method for computing R_0^d method can be generalized to capture all four pair types at once, but in this case, due to the increased dimensionality of the transition matrix, calculations will become less transparent. It is worth noting that when analytic calculation may fail to give a closed form expression, this technique can be easily implemented using a simple programming language to derive numerical estimates and to assess the implications of truncating the recurrence equations at some finite generation.

5. Discussion and further work

Incorporating behavioural change into epidemiological models is a challenging task with many unknowns when modelling the transmission of information and responsiveness of people. These are complex processes with many heterogeneities at the individual level in how information is acquired, processed, acted upon and transmitted further. In this paper, we derived and analyzed a pairwise and simulation model that capture multiple ways of generating and transmitting information as well as the overlap between routes of disease and information transmission. The pairwise model, for two completely overlapping networks, was used to assess the efficacy of different sources of information generation and routes of information transmission in bringing prevalence to as a low level as possible. Contact based transmission of information was found to be the most efficient as it creates multiple secondary sources of information and outcompetes processes such as the global transmission of information. The analysis, in line with previous findings [11], also shows that information cannot always stop an epidemic but can significantly reduce its impact via lowering the prevalence of infection.

The individual based model was used to study the effects of network structure using a simple similarity measure, the neighborhood overlap of an individual between disease and information transmission (i.e. G_d and G_i). We showed evidence to suggest that the effect of neighborhood overlap is tightly linked to the primary generation of information and can result in the creation of different correlation structures in responsiveness (see Fig. 8) and also between responsiveness and infection (see Fig. 4). The particular types of information generation and transmission will heavily depend on disease characteristics and these will impact on model outcome and on establishing what the most influential factors are in controlling epidemics via information transmission. A more realistic specification of the information and infection networks aided by real world data, for example from sexual health surveys and email exchange volumes, would allows us to further tease apart the importance of edge overlap as well as identify other network measures that may be significant, such as the path length difference between networks and the occurrence of network motifs, and examine their effect on the two diffusion processes. It is rarely the case that diseases invade populations that are fully naive. A number of responsive individuals are always present due to past epidemics or simply due to awareness being a trait that arises independently of past epidemics through genetic, social, cultural or economic reasons. In either case, the number of responsive individuals and their precise distribution will have an important impact on whether newly seeded infections can invade. In this paper, we presented a semi-analytic calculation of R_0 based on probabilistic arguments and showed that the mixing patterns in information can significantly alter the basic reproductive number of the disease. This is an important result since it opens up the possibility to tune or optimize responsiveness in order to limit the potential of new outbreaks.

A number of simplifying assumptions on human behavior are built into our model. There are only two levels of information awareness and transitions between them are instantaneous. In the real world an individual's exposure to the risk of infection is certainly not limited to two categories. People are also likely to react to information around them via imitative behavior with individuals having different propensity for changing their behavior. Furthermore the decision to act on information may depend on the source of information - or the combination of sources of information - the order of arrival of information, or any segmentation of the population as discussed in the introduction. Our individual based model is designed such that it can accommodate further behavioral heterogeneities, such as those mentioned above. There is now enough evidence to suggest that cognitive and social aspects of a host population, as well as a network of contacts (particularly in the case of

STD's), are interacting and altering the epidemiological profile of an infectious pathogen. These are early days for behaviorial epidemiology and in the future it is likely that knowledge generated in other fields, for example cognitive and evolutionary psychology, communications theory, market research, game theory, anthropology, geography and demography should be examined for integration into epidemiological modelling.

Acknowledgements

Vasilis Hatzopoulos and Istvan Z. Kiss acknowledge support from EPSRC (EP/H001085/1). Michael Taylor acknowledges support from EPSRC (DTA grant). Péter L. Simon acknowledges support from OTKA (Grant No.81403).

Appendix A. Pair-approximation equations

Here, the equation of the pair-approximation model that accounts for disease transmission through a static network of contacts and three different routes of responsiveness transmission are given. The first route of responsiveness transmission overlaps completely with the disease transmission route, while the second and third account for mean-field and global transmission of information, respectively. The equations are:

$$\begin{aligned}[\dot{S_{nr}}] = &-\tau[S_{nr}I_{nr}] - \tau\sigma_i[S_{nr}I_r] + pr[T] - \lambda_C\alpha_s([S_{nr}S_r] + [S_{nr}I_r] \\ &+ [S_{nr}T]) - \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}]/N \\ &- \lambda_G G_s([I_{nr}],[I_r])[S_{nr}] + d_s[S_r],\end{aligned} \tag{15}$$

$$\begin{aligned}[\dot{S_r}] = &-\tau\sigma_s[S_rI_{nr}] - \tau\sigma_i\sigma_s[S_rI_r] + (1-p)r[T] + \lambda_C\alpha_s([S_{nr}S_r] \\ &+ [S_{nr}I_r] + [S_{nr}T]) + \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}]/N \\ &+ \lambda_G G_s([I_{nr}],[I_r])[S_{nr}] - d_s[S_r],\end{aligned} \tag{16}$$

$$\begin{aligned}[\dot{I_{nr}}] = &+\tau[S_{nr}I_{nr}] + \tau\sigma_i[S_{nr}I_r] - \gamma[I_{nr}] - \lambda_C\alpha_i([I_{nr}S_r] + [I_{nr}I_r] \\ &+ [I_{nr}T]) - \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}]/N \\ &- \lambda_G G_i([I_{nr}],[I_r])[I_{nr}] + d_i[I_r] - \omega[I_{nr}],\end{aligned} \tag{17}$$

$$\begin{aligned}[\dot{I_r}] = &+\tau\sigma_s[S_rI_{nr}] + \tau\sigma_i\sigma_s[S_rI_r] - \gamma\sigma_r[I_r] + \lambda_C\alpha_i([I_{nr}S_r] + [I_{nr}I_r] \\ &+ [I_{nr}T]) + \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}]/N \\ &+ \lambda_G G_i([I_{nr}],[I_r])[I_{nr}] - d_i[I_r] + \omega[I_{nr}],\end{aligned} \tag{18}$$

$$[\dot{T}] = +\gamma[I_{nr}] + \gamma\sigma_r[I_r] - r[T], \tag{19}$$

$$\begin{aligned}[\dot{S_{nr}S_{nr}}] = &-2\tau[S_{nr}S_{nr}I_{nr}] - 2\tau\sigma_i[S_{nr}S_{nr}I_r] + 2pr[S_{nr}T] \\ &+ 2d_s[S_{nr}S_r] - 2\lambda_C\alpha_s([S_{nr}S_{nr}S_r] + [S_{nr}S_{nr}I_r] \\ &+ [S_{nr}S_{nr}T]) \ - 2\lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}S_{nr}]/N \\ &- 2\lambda_G G_s([I_{nr}],[I_r])[S_{nr}S_{nr}],\end{aligned} \tag{20}$$

$$\begin{aligned}[\dot{S_{nr}S_r}] = &-\tau\sigma_s[S_{nr}S_rI_{nr}] - \tau\sigma_i\sigma_s[S_{nr}S_rI_r] - \tau[I_{nr}S_{nr}S_r] \\ &- \tau\sigma_i[I_rS_{nr}S_r] + pr[TS_r] + (1-p)r[S_{nr}T] \\ &+ \lambda_C\alpha_s([S_{nr}S_{nr}S_r] + [S_{nr}S_{nr}I_r] + [S_{nr}S_{nr}T]) \\ &+ \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}S_{nr}]/N \\ &+ \lambda_G G_s([I_{nr}],[I_r])[S_{nr}S_{nr}] - \lambda_C\alpha_s([S_rS_{nr}S_r] + [I_rS_{nr}S_r] \\ &+ [TS_{nr}S_r]) - \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}S_r]/N \\ &- \lambda_G G_s([I_{nr}],[I_r])[S_{nr}S_r] - d_s[S_{nr}S_r] + d_s[S_rS_r] \\ &- \lambda_C\alpha_s[S_{nr}S_r],\end{aligned} \tag{21}$$

$$\begin{aligned}[\dot{S_rS_r}] = &-2\tau\sigma_s[S_rS_rI_{nr}] - 2\tau\sigma_s\sigma_i[S_rS_rI_r] + 2(1-p)r[S_rT] \\ &+ 2\lambda_C\alpha_s([S_rS_{nr}S_r] + [I_rS_{nr}S_r] + [TS_{nr}S_r]) \\ &+ 2\lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}S_r]/N \\ &+ 2\lambda_G G_s([I_{nr}],[I_r])[S_{nr}S_r] - 2d_s[S_rS_r] + 2\lambda_C\alpha_s[S_{nr}S_r],\end{aligned} \tag{22}$$

$$\begin{aligned}[\dot{S_{nr}I_{nr}}] = &+\tau[S_{nr}S_{nr}I_{nr}] + \tau\sigma_i[S_{nr}S_{nr}I_r] - \tau[I_{nr}S_{nr}I_{nr}] \\ &- \tau\sigma_i[I_rS_{nr}I_{nr}] - \tau[S_{nr}I_{nr}] - \lambda_C\alpha_s([S_rS_{nr}I_{nr}] + [I_rS_{nr}I_{nr}] \\ &+ [TS_{nr}I_{nr}]) - \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}I_{nr}]/N \\ &- \lambda_G G_s([I_{nr}],[I_r])[S_{nr}I_{nr}] - \lambda_C\alpha_i([S_{nr}I_{nr}S_r] + [S_{nr}I_{nr}I_r] \\ &+ [S_{nr}I_{nr}T]) - \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[S_{nr}I_{nr}]/N \\ &- \lambda_G G_i([I_{nr}],[I_r])[S_{nr}I_{nr}] - \gamma[S_{nr}I_{nr}] + rp[TI_{nr}] + d_i[S_{nr}I_r] \\ &+ d_s[S_rI_{nr}] - \omega[S_{nr}I_{nr}],\end{aligned} \tag{23}$$

$$\begin{aligned}[\dot{S_{nr}I_r}] = &+\tau\sigma_s[S_{nr}S_rI_{nr}] + \tau\sigma_i\sigma_s[S_{nr}S_rI_r] - \tau[I_{nr}S_{nr}I_r] \\ &- \tau\sigma_i[I_rS_{nr}I_r] - \tau\sigma_i[S_{nr}I_r] + \lambda_C\alpha_i([S_{nr}I_{nr}S_r] + [S_{nr}I_{nr}I_r] \\ &+ [S_{nr}I_{nr}T]) + \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[S_{nr}I_{nr}]/N \\ &+ \lambda_G G_i([I_{nr}],[I_r])[S_{nr}I_{nr}] - \lambda_C\alpha_s([S_rS_{nr}I_r] + [I_rS_{nr}I_r] \\ &+ [TS_{nr}I_r]) - \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}I_{nr}]/N \\ &- \lambda_G G_s([I_{nr}],[I_r])[S_{nr}I_{nr}] - \gamma\sigma_r[S_{nr}I_r] - d_i[S_{nr}I_r] + d_s[S_rI_r] \\ &- \lambda_C\alpha_s[S_{nr}I_r] + pr[TI_r] + \omega[S_{nr}I_{nr}],\end{aligned} \tag{24}$$

$$\begin{aligned}[\dot{S_rI_{nr}}] = &+\tau[S_rS_{nr}I_{nr}] + \tau\sigma_i[S_rS_{nr}I_r] - \tau\sigma_s[I_{nr}S_rI_{nr}] \\ &- \tau\sigma_i\sigma_s[I_rS_rI_{nr}] - \tau\sigma_s[S_rI_{nr}] + \lambda_C\alpha_s([S_rS_{nr}I_{nr}] + [I_rS_{nr}I_{nr}] \\ &+ [TS_{nr}I_{nr}]) + \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}I_{nr}]/N \\ &+ \lambda_G G_s([I_{nr}],[I_r])[S_{nr}I_{nr}] - \lambda_C\alpha_i([S_rI_{nr}S_r] + [S_rI_{nr}I_r] \\ &+ [S_rI_{nr}T]) - \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[S_rI_{nr}]/N \\ &- \lambda_G G_i([I_{nr}],[I_r])[S_rI_{nr}] - \gamma[S_rI_{nr}] + (1-p)r[TI_{nr}] \\ &- d_s[S_rI_{nr}] + d_i[S_rI_r] - \lambda_C\alpha_i[S_rI_{nr}] - \omega[S_rI_{nr}],\end{aligned} \tag{25}$$

$$\begin{aligned}[\dot{S_rI_r}] = &+\tau\sigma_s[S_rS_rI_{nr}] + \tau\sigma_i\sigma_s[S_rS_rI_r] - \tau\sigma_s[I_{nr}S_rI_r] \\ &- \tau\sigma_i\sigma_s[I_rS_rI_r] - \tau\sigma_i\sigma_s[S_rI_r] + \lambda_C\alpha_s([S_rS_{nr}I_r] + [I_rS_{nr}I_r] \\ &+ [TS_{nr}I_r]) + \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}I_r]/N \\ &+ \lambda_G G_s([I_{nr}],[I_r])[S_{nr}I_r] + \lambda_C\alpha_i([S_rI_{nr}S_r] + [S_rI_{nr}I_r] \\ &+ [S_rI_{nr}T]) + \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[S_rI_{nr}]/N \\ &+ \lambda_G G_i([I_{nr}],[I_r])[S_rI_{nr}] - \gamma\sigma_r[S_rI_r] + (1-p)r[TI_r] \\ &- d_s[S_rI_r] - d_i[S_rI_r] + \lambda_C\alpha_s[S_{nr}I_r] + \lambda_C\alpha_i[S_rI_{nr}] + \omega[S_rI_{nr}],\end{aligned} \tag{26}$$

$$\begin{aligned}[\dot{I_{nr}I_{nr}}] = &+2\tau[I_{nr}S_{nr}I_{nr}] + 2\tau\sigma_i[I_{nr}S_{nr}I_r] + 2\tau[S_{nr}I_{nr}] - 2\gamma[I_{nr}I_{nr}] \\ &- 2\lambda_C\alpha_i([I_{nr}I_{nr}S_r] + [I_{nr}I_{nr}I_r] + [I_{nr}I_{nr}T]) \\ &- 2\lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}I_{nr}]/N \\ &- 2\lambda_G G_i([I_{nr}],[I_r])[I_{nr}I_{nr}] + 2d_i[I_{nr}I_r] - 2\omega[I_{nr}I_{nr}],\end{aligned} \tag{27}$$

$$\begin{aligned}[\dot{I_{nr}I_r}] = &+\tau[I_{nr}S_{nr}I_r] + \tau\sigma_i[I_rS_{nr}I_r] + \tau\sigma_s[I_{nr}S_rI_{nr}] + \tau\sigma_i\sigma_s[I_{nr}S_rI_r] \\ &+ \tau\sigma_i[S_{nr}I_r] + \tau\sigma_s[I_{nr}S_r] - (\gamma+\gamma\sigma_r)[I_{nr}I_r] \\ &+ \lambda_C\alpha_i([I_{nr}I_{nr}S_r] + [I_{nr}I_{nr}I_r] + [I_{nr}I_{nr}T]) + \lambda_{MF}m_ik_{MF}([S_r] \\ &+ [I_r] + [T])[I_{nr}I_{nr}]/N + \lambda_G G_i([I_{nr}],[I_r])[I_{nr}I_{nr}] \\ &- \lambda_C\alpha_i([S_rI_{nr}I_r] + [I_rI_{nr}I_r] + [TI_{nr}I_r]) - \lambda_{MF}m_ik_{MF}([S_r] \\ &+ [I_r] + [T])[I_{nr}I_r]/N - \lambda_G G_i([I_{nr}],[I_r])[I_{nr}I_r] - d_i[I_{nr}I_r] \\ &+ d_i[I_rI_r] - \lambda_C\alpha_i[I_{nr}I_r] + \omega[I_{nr}I_{nr}] - \omega[I_{nr}I_r],\end{aligned} \tag{28}$$

$$[I_r\dot{I}_r] = +2\tau\sigma_s[I_{nr}S_rI_{nr}] + 2\tau\sigma_i\sigma_s[I_rS_rI_r] + 2\tau\sigma_i\sigma_s[S_rI_r] + 2\lambda_C\alpha_i([S_rI_{nr}I_r] + [I_rI_{nr}I_r] + [TI_{nr}I_r]) + 2\lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}I_r]/N + 2\lambda_GG_i([I_{nr}],[I_r])[I_{nr}I_r] - 2\gamma\sigma_r[I_rI_r] + 2\lambda_C\alpha_i[I_rI_{nr}] - 2d_i[I_rI_r] + 2\omega[I_{nr}I_r], \quad (29)$$

$$[S_{nr}\dot{T}] = -\tau[I_{nr}S_{nr}T] - \tau\sigma_i[I_rS_{nr}T] - \lambda_C\alpha_s([S_rS_{nr}T] + [I_rS_{nr}T] + [TS_{nr}T]) - \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}T]/N - \lambda_GG_s([I_{nr}],[I_r])[S_{nr}T] + \gamma[S_{nr}I_{nr}] + \gamma\sigma_r[S_{nr}I_r] + d_s[S_rT] + pr[TT] - r[S_{nr}T] - \lambda_C\alpha_s[S_{nr}T], \quad (30)$$

$$[S_r\dot{T}] = -\tau\sigma_s[I_{nr}S_rT] - \tau\sigma_i\sigma_s[I_rS_rT] + \lambda_C\alpha_s([S_rS_{nr}T] + [I_rS_{nr}T] + [TS_{nr}T]) + \lambda_{MF}m_sk_{MF}([S_r] + [I_r] + [T])[S_{nr}T]/N + \lambda_GG_s([I_{nr}],[I_r])[S_{nr}T] + \gamma[S_rI_{nr}] + \gamma\sigma_r[S_rI_r] - d_s[S_rT] - r[S_rT] + r(1-p)[TT] + \lambda_C\alpha_s[S_{nr}T], \quad (31)$$

$$[I_{nr}\dot{T}] = +\tau[I_{nr}S_{nr}T] + \tau\sigma_i[I_rS_{nr}T] - \lambda_C\alpha_i([S_rI_{nr}T] + [I_rI_{nr}T] + [TI_{nr}T]) - \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}T]/N - \lambda_GG_i([I_{nr}],[I_r])[I_{nr}T] + \gamma[I_{nr}I_{nr}] + \gamma\sigma_r[I_{nr}I_r] + d_i[I_rT] - r[I_{nr}T] - \gamma[I_{nr}T] - \lambda_C\alpha_i[I_{nr}T] - \omega[I_{nr}T], \quad (32)$$

$$[I_r\dot{T}] = +\tau\sigma_s[I_{nr}S_rT] + \tau\sigma_i\sigma_s[I_rS_rT] + \lambda_C\alpha_i([S_rI_{nr}T] + [I_rI_{nr}T] + [TI_{nr}T]) + \lambda_{MF}m_ik_{MF}([S_r] + [I_r] + [T])[I_{nr}T]/N + \lambda_GG_i([I_{nr}],[I_r])[I_{nr}T] + \gamma[I_rI_{nr}] + \gamma\sigma_r[I_rI_r] - \gamma\sigma_r[I_rT] - r[I_rT] + \lambda_C\alpha_i[I_{nr}T] - d_i[I_rT] + \omega[I_{nr}T], \quad (33)$$

$$[T\dot{T}] = 2\gamma[I_{nr}T] + 2\gamma\sigma_r[I_rT] - 2r[TT]. \quad (34)$$

where $N = S_{nr} + S_r + I_{nr} + I_r + T$ is the population size. The mean-field type transmission is implemented based on ideas proposed by Eames [6]. The routes of responsiveness transmission can be switched on and off or can be combined using λ_C,λ_{MF} and λ_G which are set to 0 or 1 accordingly. For our purposes we set

$$G_s([I_{nr}],[I_r]) = G_i([I_{nr}],[I_r]) = \frac{\delta([I_{nr}]+[I_r])^n}{K + ([I_{nr}]+[I_r])^n}. \quad (35)$$

In this paper, $n = 1$ at all times. To integrate the equations numerically, we use the classic closure proposed in [18]. This amounts to approximating all triples in terms of singles and pairs with the general closure relation given by

$$[ABC] = \frac{\langle k\rangle - 1}{\langle k\rangle}\frac{[AB][BC]}{[B]}.$$

This approximation closes the system and numerical integration can be performed. Parameter values are based on those used in [20] but with units changed from weeks to years. For simplicity we assume that $\alpha_s = \alpha_i = \alpha$, $d_s = d_i = d$ and $\delta_s = \delta_i = \delta$.

Appendix B. List of all possible transitions and their probability

From Figs. 5 and 6(a), it is easy to note that only the following transitions are possible,

$$A \to B, B \to \{A, D\}, C \to \{A, D\} \text{ and } D \to \{B, C\}, \quad (36)$$

with the probability of transitions given by

$$r_{AB} = \frac{\omega}{\gamma_{nr} + \omega + \tau}, \quad (37)$$

$$r_{BA} = \frac{d}{\tau\sigma_i + \alpha + d + \gamma_r}, r_{BD} = \frac{\alpha}{\tau\sigma_i + \alpha + d + \gamma_r}, \quad (38)$$

$$r_{CA} = \frac{d}{\gamma_{nr} + \omega + \alpha + \tau\sigma_s + d}, r_{CD} = \frac{\alpha + \omega}{\gamma_{nr} + \omega + \alpha + \tau\sigma_s + d}, \quad (39)$$

$$r_{DB} = \frac{d}{\gamma_r + 2d + \tau\sigma_s\sigma_i}, r_{DC} = \frac{d}{\gamma_r + 2d + \tau\sigma_s\sigma_i}. \quad (40)$$

Appendix C. Solving linear recurrence equations

Simultaneous recurrence equations can be solved using simple linear algebra arguments. Below we give a general result for k equations followed by the full analytic solution for two equations as needed for the R_0^d calculations.

Proposition 1. *Let A be a $k \times k$ diagonalisable matrix. Hence, there is an invertible matrix P, such that*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_k \end{pmatrix},$$

where λ_i, $i = 1, 2, \ldots k$ are the eigenvalues of A, and assume that $|\lambda_i| < 1$ for all $i = 1, 2, \ldots k$. Let $x(n)$ be defined by

$$x(n+1) = Ax(n), \quad x(0) = x_0 \in \mathbb{R}^k.$$

Then

$$\sum_{n=0}^{\infty} x(n) = PDP^{-1}x_0, \quad (41)$$

where D is a diagonal matrix with diagonal elements $D_{ii} = \frac{1}{1-\lambda_i}$ for all $i = 1, 2, \ldots k$.

Proof. It follows immediately that $(P^{-1}AP)^n = P^{-1}A^nP$. Hence, $A^n = P(P^{-1}AP)^nP^{-1} = P\Lambda_nP^{-1}$, where Λ_n is a diagonal matrix with diagonal elements $\Lambda_{ii} = \lambda_i^n$, where $i = 1, 2, \ldots k$ and $n = 0, 1, 2, \ldots$. Therefore, $\sum_{n=0}^{\infty}\Lambda_n = D$. Thus we obtain

$$\sum_{n=0}^{\infty} x(n) = \sum_{n=0}^{\infty} A^nx_0 = \sum_{n=0}^{\infty} P\Lambda_nP^{-1}x_0 = P\left(\sum_{n=0}^{\infty}\Lambda_n\right)P^{-1}x_0 = PDP^{-1}x_0. \quad \square$$

Let us now apply the above result for a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The following two statements can by verified by elementary calculations. The matrix is diagonalisable if and only if

$$(a_{11} - a_{22})^2 + 4a_{12}a_{21} = 0 \text{ implies } a_{12} = a_{21} = 0. \quad (42)$$

The eigenvalues of matrix A are within the unit circle (i.e. $|\lambda_i| < 1$ for $i = 1,2$) if and only if

$$|\mathrm{Tr}A| - 1 < \det A < 1. \quad (43)$$

Hence the following proposition holds,

Proposition 2. *For a 2×2 matrix A which satisfies conditions (42) and (43), the following equation holds,*

$$\sum_{n=0}^{\infty} x(n) = \frac{1}{1 - \mathrm{Tr}A + \det A}\begin{pmatrix} 1 - a_{22} & a_{12} \\ a_{21} & 1 - a_{11} \end{pmatrix}x_0. \quad (44)$$

Proof. Let us denote the eigenvalues of A by λ_1, λ_2 and the eigenvectors by u and v. Then P contains u and v as column vectors: $P = (u\ v)$, and the eigenvectors can be expressed as

$$u = (a_{12}, \lambda_1 - a_{11})^T, \quad v = (a_{12}, \lambda_2 - a_{11})^T.$$

Then

$$P = \begin{pmatrix} a_{12} & a_{12} \\ \lambda_1 - a_{11} & \lambda_2 - a_{11} \end{pmatrix}, \quad P^{-1} = \frac{1}{a_{12}(\lambda_2 - \lambda_1)} \begin{pmatrix} \lambda_2 - a_{11} & -a_{12} \\ -\lambda_1 + a_{11} & a_{12} \end{pmatrix}.$$

After some algebra, and using that $\lambda_1 + \lambda_2 = \mathrm{Tr}A$ and $\lambda_1\lambda_2 = \det A$, the following identity holds,

$$PDP^{-1} = \frac{1}{1 - \mathrm{Tr}A + \det A} \begin{pmatrix} 1 - a_{22} & a_{12} \\ a_{21} & 1 - a_{11} \end{pmatrix}. \qquad \square$$

Using the above Proposition in the case of $x^0 = (x_1^0, x_2^0)^T$, the first and second coordinate of $x(n)$ are given by

$$\sum_{n=0}^{\infty} x_1(n) = \frac{(1 - a_{22})x_1^0 + a_{12}x_2^0}{1 - \mathrm{Tr}A + \det A}, \tag{45}$$

$$\sum_{n=0}^{\infty} x_2(n) = \frac{a_{21}x_1^0 + (1 - a_{11})x_2^0}{1 - \mathrm{Tr}A + \det A}. \tag{46}$$

Hence, for $x_0 = (1,0)^T$ and $x_0 = (0,1)^T$, the first and second coordinate of $x(n)$ are given by

$$\sum_{n=0}^{\infty} x_1(n) = \frac{1 - a_{22}}{1 - \mathrm{Tr}A + \det A}, \tag{47}$$

$$\sum_{n=0}^{\infty} x_2(n) = \frac{a_{21}}{1 - \mathrm{Tr}A + \det A} \tag{48}$$

and

$$\sum_{n=0}^{\infty} x_1(n) = \frac{a_{12}}{1 - \mathrm{Tr}A + \det A}, \tag{49}$$

$$\sum_{n=0}^{\infty} x_2(n) = \frac{1 - a_{11}}{1 - \mathrm{Tr}A + \det A}. \tag{50}$$

Appendix D. Calculations for trees starting with pair (S_{nr},I_r), (S_r,I_{nr}) and (S_r,I_r)

Here the same techniques that were used to derive the tree for (S_{nr},I_{nr}) are used for the other three combinations of S and I individuals.

D.1. Tree starting with pair (S_{nr},I_r)

This tree is easy to create and the same observations as before apply. In particular, B and/or C only appear in every even generation while A and D are present in every odd one. Hence, the coupled recurrence equations are given by

$$b^B(n+1) = r_{BB}^B b^B(n) + r_{CB}^B c^B(n), \tag{51}$$

$$c^B(n+1) = r_{BC}^B b^B(n) + r_{CC}^B c^B(n), \tag{52}$$

with initial condition $(b^B(0),\ c^B(0)) = (1,0)$, where $r_{BB}^B = r_{BA}r_{AB} + r_{BD}r_{DB}, r_{BC}^B = r_{BD}r_{DC}, r_{CB}^B = r_{CA}r_{AB} + r_{CD}r_{DB}$ and $r_{CC}^B = r_{CD}r_{DC}$. Using Eqs. (47) and (48) developed in Appendix C gives

$$P_B^B = \frac{1 - r_{CC}^B}{1 - (r_{BB}^B + r_{CC}^B) + r_{BB}^B r_{CC}^B - r_{BC}^B r_{CB}^B},$$

$$P_C^B = \frac{r_{BC}^B}{1 - (r_{BB}^B + r_{CC}^B) + r_{BB}^B r_{CC}^B - r_{BC}^B r_{CB}^B}. \tag{53}$$

Given that each B generates an A and a D with probability r_{BA} and r_{BD}, and that each C generates an A and two Ds with probability r_{CA} and r_{CD}, it follows that

$$P_A^B = P_B^B r_{BA} + P_C^B r_{CA}, P_D^B = P_B^B r_{BD} + P_C^B r_{CD}. \tag{54}$$

Therefore the probability of an infection from an (S_{nr},I_r) pair is

$$P((S_{nr}, I_r)) = \sum_{X \in \{A,B,C,D\}} p_X^{inf} P_X^B. \tag{55}$$

D.2. Tree starting with pair (S_r,I_{nr})

Working out $P((S_r,I_{nr}))$ follows a similar line of thought and starts by generating a tree beginning with C. Exactly the same observations as above hold (i.e. B and/or C only appear in every even generation while A and D are present in every odd one). Using the same approach as above the coupled recurrence equations now read as

$$b^C(n+1) = r_{BB}^C b^C(n) + r_{CB}^C c^C(n), \tag{56}$$

$$c^C(n+1) = r_{BC}^C b^C(n) + r_{CC}^C c^C(n), \tag{57}$$

with initial condition $(b^C(0),c^C(0)) = (0,1)$, where $r_{BB}^C = r_{BA}r_{AB} + r_{BD}r_{DB}, r_{BC}^C = r_{BD}r_{DC}, r_{CB}^C = r_{CA}r_{AB} + r_{CD}r_{DB}$ and $r_{CC}^C = r_{CD}r_{DC}$. Using Eqs. (49) and (50) developed in Appendix C gives

$$P_B^C = \frac{r_{CB}^C}{1 - (r_{BB}^C + r_{CC}^C) + r_{BB}^C r_{CC}^C - r_{BC}^C r_{CB}^C},$$

$$P_C^C = \frac{1 - r_{BB}^C}{1 - (r_{BB}^C + r_{CC}^C) + r_{BB}^C r_{CC}^C - r_{BC}^C r_{CB}^C}. \tag{58}$$

Given that each B generates an A and a D with probability r_{BA} and r_{BD}, and that each C generates an A and two Ds with probability r_{CA} and r_{CD}, it follows that

$$P_A^C = P_B^C r_{BA} + P_C^C r_{CA}, P_D^C = P_B^C r_{BD} + P_C^C r_{CD}. \tag{59}$$

Therefore the probability of an infection from an (S_r,I_{nr}) pair is

$$P((S_r, I_{nr})) = \sum_{X \in \{A,B,C,D\}} p_X^{inf} P_X^C. \tag{60}$$

D.3. Tree starting with pair (S_r,I_r)

The calculations $P((S_r,I_r))$ are identical and rely on the same technique of generating a new tree starting with pair D. Once this is done, the coupled recurrence equations are given by

$$a^D(n+1) = r_{AA}^D a^D(n) + r_{DA}^D d^D(n), \tag{61}$$

$$d^D(n+1) = r_{AD}^D a^D(n) + r_{DD}^D d^D(n), \tag{62}$$

with initial condition $(a^D(0),d^D(0)) = (0,1)$, where $r_{AA}^D = r_{AB}r_{BA}, r_{AD}^D = r_{AB}r_{BD}, r_{DA}^D = r_{DB}r_{BA} + r_{DC}r_{CA}$ and $r_{DD}^D = r_{DB}r_{BD} + r_{DC}r_{CD}$. Using Eqs. (49) and (50) developed in Appendix C gives

$$P_A^D = \frac{r_{DA}^D}{1 - (r_{AA}^D + r_{DD}^D) + r_{AA}^D r_{DD}^D - r_{AD}^D r_{DA}^D},$$

$$P_D^D = \frac{1 - r_{AA}^D}{1 - (r_{AA}^D + r_{DD}^D) + r_{AA}^D r_{DD}^D - r_{AD}^D r_{DA}^D}. \tag{63}$$

Given that each A generates a B with probability r_{AB} and that each D generates a B and a C with probability r_{DB} and r_{DC}, it follows that

$$P_B^D = P_A^D r_{AB} + P_D^D r_{DB}, P_C^D = P_D^D r_{DC}. \tag{64}$$

Therefore the probability of an infection from an (S_r,I_r) pair is

$$P((S_r, I_r)) = \sum_{X \in \{A,B,C,D\}} p_X^{inf} P_X^D. \tag{65}$$

References

[1] C.T. Bauch, Imitation dynamics predicting vaccinating behaviour, Proc. R. Soc. B. 272 (2005) 1669.
[2] C.T. Bauch, D.J.D. Earn, Vaccination and the theory of games, PNAS 101 (36) (2004) 13391.
[3] M.H. Becker, J.G. Joseph, Aids and behavioral change to reduce risk:a review, Am. J. Public Health 78 (4) (1988) 395.
[4] J.T. Bertrand et al., Systematic review of the effectiveness of mass communication programs to change hiv/aids-related behaviors in developing countries, Health Educ. Res. 21 (4) (2006) 567.
[5] A. d'Onofrio, P. Manfredi, E. Salinelli, Vaccinating behaviour, information, and the dynamics of sir vaccine preventable diseases, Theor. Pop. Biol. 71 (2007) 301.
[6] K.T.D. Eames, Modelling disease spread through random and regular contacts in clustered populations, Theor. Popul. Biol. 73 (2008) 104.
[7] K.T.D. Eames, M.J. Keeling, Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases, Proc. Natl. Acad. Sci. USA. 99 (20) (2002) 13330.
[8] J.M. Esptein, J. Parker, D. Cummings, R.A. Hammond, Coupled contagion dynamics of fear and disease: mathematical and computational explorations, PLOS ONE 3 (12) (2008) e3955.
[9] N. Ferguson, Capturing human behavior, Nature 446 (2007) 733.
[10] S. Funk, E. Gilad, V.A.A. Jansen, Endemic disease, awareness and local behavioural response, J. Theor. Biol. 264 (2) (2010) 501.
[11] S. Funk, E. Gilad, C. Watkins, V.A.A. Jansen, The spread of awareness and its impact on epidemic outbreaks, PNAS 106 (16) (2009) 6872.
[12] S. Funk, M. Salathe, V.A.A. Jansen, Modelling the influence of human behavior on the spread of infectious diseases: a review, J. R. Soc. Inter. 106 (16) (2010) 6872.
[13] J. Goldenberg, B. Libai, E. Muller, Talk of the network: a complex-systems look at the underlying process of word-of-mouth, Market Lett. 12 (3) (2001) 211.
[14] G. Gregson et al., Hiv decline associated with behavior change in eastern zimbabwe, Science 311 (5761) (2006) 664.
[15] D.T. Hamilton, M.S. Handcock, M. Morris, Degree distributions in sexual networks: a framework for evaluating evidence, Sex. Transm. Dis 35 (1) (2008) 30.
[16] T. House, M.J. Keeling, The impact of contact tracing in clustered populations, PLoS Comput. Biol. 266 (2010) e1000721.
[17] S.C. Kalichman, L.C. Simbayi, Hiv testing attitudes, aids stigma, and voluntary hiv counselling and testing in a black township in cape town, south africa, Sex. Transm. Infect. 79 (2003) 442.
[18] M.J. Keeling, The effects of local spatial structure on epidemiological invasions, Proc. R. Soc. Lond. B 266 (1999) 859.
[19] I.Z. Kiss, M. Broom, P.G. Craze, Can epidemic models describe the diffusion of topics across disciplines?, J Informetr. 4 (2010) 74.
[20] I.Z. Kiss, J. Cassell, M. Recker, P.L. Simon, The effect of information transmission on epidemic outbreaks, Math. Biosci. 225 (1) (2009) 1.
[21] A. Kleczkowski, S. Maharaj, Stay at home, wash your hands: Epidemic dynamics with awareness of infection. 446 (2007) 733.
[22] J. Kleinberg, Cascading Behavior in Networks: Algorithmic and Economic Issues, Cambridge University Press, 2007. chapter 24.
[23] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, ACM Trans.Web 1 (1) (2007) 5.
[24] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aberg, The web of human sexual contacts, Nature 411 (2001) 907.
[25] N. Meade, T. Islam, Modelling and forecasting the duffusion of innovation - a 25-year review, Int. J. Forecast. 22 (2006) 519.
[26] A. Perisic, C.T. Bauch, Social contact networks and disease eradicability under voluntary vaccination, Plos. Comput. Biol. 5 (2) (2009) e1000280, doi:10.1371/journal.pcbi.1000280.
[27] J.E. Phelps, R. Lewis, L. Mobilio, D. Perry, N. Raman, Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email, J. Advertis. res. 44 (4) (2005) 333.
[28] P. Poletti, B. Caprile, M. Ajelli, A. Pugliese, S. Merler, Spontaneous behavioural changes in response to epidemics, J. Theor. Biol. 260 (2009) 31.
[29] J. J Potterat, S.Q. Muth, P.B. Rothenberg, H. Zimmerman-Rogers, D.L. Green, J.E. Taylor, M.S. Bonney, H.A. White, Sexual network structure as an indicator of epidemic phase, Sex. Transm. Infect 78 (2002) i152.
[30] T. Rehle et al., National hiv incidence measures-new insights into the south african epidemic, S. Afr. Med. J. 97 (2007) 194.
[31] T.W. Valente, R. Fosados, Diffusion of innovations and network segmentation: the part played by people in promoting health, Sex. Transm. Dis 33 (7) (2006) s23.