

Incorporating Disease and Population Structure into Models of SIR Disease in Contact Networks

Joel C. Miller^{1*}, Erik M. Volz²

1 Departments of Mathematics and Biology, Penn State University, University Park, Pennsylvania, United States of America, **2** Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, United States of America

Abstract

We consider the recently introduced edge-based compartmental models (EBCM) for the spread of susceptible-infected-recovered (SIR) diseases in networks. These models differ from standard infectious disease models by focusing on the status of a random partner in the population, rather than a random individual. This change in focus leads to simple analytic models for the spread of SIR diseases in random networks with heterogeneous degree. In this paper we extend this approach to handle deviations of the disease or population from the simplistic assumptions of earlier work. We allow the population to have structure due to effects such as demographic features or multiple types of risk behavior. We allow the disease to have more complicated natural history. Although we introduce these modifications in the static network context, it is straightforward to incorporate them into dynamic network models. We also consider serosorting, which requires using dynamic network models. The basic methods we use to derive these generalizations are widely applicable, and so it is straightforward to introduce many other generalizations not considered here. Our goal is twofold: to provide a number of examples generalizing the EBCM method for various different population or disease structures and to provide insight into how to derive such a model under new sets of assumptions.

Citation: Miller JC, Volz EM (2013) Incorporating Disease and Population Structure into Models of SIR Disease in Contact Networks. PLoS ONE 8(8): e69162. doi:10.1371/journal.pone.0069162

Editor: Alain Barrat, Centre de Physique Théorique, France

Received: November 7, 2012; **Accepted:** June 10, 2013; **Published:** August 19, 2013

Copyright: © 2013 Miller, Volz. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JCM was supported in part by 1) the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health and 2) the Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health under Award Number U54GM088558 from the National Institute Of General Medical Sciences. EMV was supported by NIH K01 AI091440. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: joel.c.miller.research@gmail.com

Introduction

The social contact network of a population plays a significant role in controlling the spread of directly transmitted infectious diseases. The accuracy of our predictions about the course of an epidemic or the effectiveness of an intervention is thus tied closely to our ability to accurately model the impact of the contact network on disease transmission. Two of the features that are most notable in network observations are that partnerships may have significant duration and different individuals have different numbers of partners. Recent work [1–3] has provided a modeling approach referred to as “edge-based compartmental modeling” (EBCM) which incorporates both partnership duration and heterogeneities in numbers of partners for susceptible-infected-recovered (SIR) epidemics. These studies find surprisingly simple equations governing the macroscopic dynamics as the disease spreads through random contact networks. The EBCM models are expected to be exact in the large-population limit when the population satisfies appropriate assumptions, and have been proved exact in the special case of static Configuration Model populations [4].

Mechanistic infection-spread models which explicitly incorporate interactions between infected and susceptible individuals typically fall into one of two classes. We will refer to the most-common of these as “mass-action” models. In the simplest mass

action model of a homogeneous population, the probability that two individuals u and v interact in a short time interval is the same for any pair of individuals, and does not depend on previous interactions. In more complex versions, we may stratify the population by demographic group (or other “type”). The probability that u and v interact in a short time interval is given by the expected number of interactions between a pair of individuals in each of their groups but is independent of whether they have interacted previously. This allows us to capture biased mixing among children and adults or simply heterogeneities in numbers of partners *e.g.*, [5–9].

An important but often unrecognized assumption of mass-action models is that they implicitly assume partnerships have zero duration, so that at any moment the partners are randomly chosen from a large pool of potential partners. We can highlight the significance of this assumption by considering a married couple. These models assume that if a man gives his wife influenza, then an arbitrarily short time later he has a new wife who may still be susceptible and his original wife has a new husband who may be susceptible. The deterministic equations that result are exact for a population that satisfies these assumptions, but the failure of the assumptions may have important implications. In real populations, we expect that over time, infected individuals tend to infect those surrounding them, leading to local depletion of susceptibles. We cannot use these models to study any effect which depends on the

duration of a partnership. For example, we cannot use mass-action models to investigate the effect of concurrency, *i.e.*, having partnerships that overlap in time, on the HIV epidemic because the mass-action equations assume that all partnerships are fleeting.

The majority of mechanistic models that are not mass action models are instead network-based. Rather than assuming partnerships have negligible duration, usually (but not always) these assume partnerships are permanent. Many of these use “pair approximations” [10–14]. Pair approximation models typically derive equations for the number of susceptible-susceptible and susceptible-infected partnerships, and often (but not always) assume that the population has homogeneous degree. The equations that arise are exact, but rely on knowledge about longer-range connections, such as the number of susceptible-susceptible-infected triples in the population. Typically an approximation is used whereby the frequency of each type of triples is expressed in terms of the pairs. For a commonly-used model network (the Configuration Model network, the simplest network model which allows for heterogeneous degree) it is possible to estimate the number of triples of various types exactly so long as the central individual is susceptible [14,15] and the initial conditions are small. If the central individual is infected or recovered, the usual estimates for the number of triples are incorrect, but these do not feed back into the equations for the pairs and so the calculations for the pairs remain correct. These models frequently require many equations, with the number of equations being $\mathcal{O}(K^2)$ where K is the number of distinct degrees in the population. Although it has recently been shown that the number of equations can be reduced dramatically without introducing errors, the approach is not obvious [14,15], and the more obvious simplifications introduce errors. With the reduction in dimension coming from [14,15] the equations reduce to the equations of the EBCM approach.

Other exact mathematical models exist that also allow for epidemics in (Configuration Model) networks having heterogeneous degree, but typically the number of equations is proportional to the maximum degree of the network or the square of the maximum degree [16,17]. If however we are willing to assume all individuals have the same number of partners, it is possible to write down a relatively simple system of equations that captures epidemic dynamics [18]. Other approaches allow us to capture the impact of dynamic partnerships, but these assume very uniform behavior across the population and occasionally assume no more than one partnership at a time [19–22].

The EBCM approach presented in [1] avoids many of the problems of other approaches, and is thus more flexible. We emphasize that this approach allows for heterogeneous degree in the underlying network. It directly leads to a simple system of equations which can be proven to exactly predict the disease dynamics in the large-population limit for a Configuration Model network [4]. The resulting system of equations has only a single governing differential equation. Subject to a few restrictions, we can adapt the assumptions so that the underlying population model differs from a Configuration Model network. The model can be adapted to include dynamic partnerships with only a small increase in complexity. Despite these advantages over existing approaches, the model retains some of the existing weaknesses. It makes very simple assumptions about the disease progression and additionally assumes that the only distinguishing feature of an individual is the number of partners. Thus it misses a large number of features: there may be a preferential direction to transmission (*e.g.*, a blood donor can infect, but not be infected by, the recipient); some individuals may be more or less susceptible to infection or more or less infectious per partnership than others (*e.g.*,

some may withdraw from society when ill while others may continue as normal); there may be biased mixing by age or by behavior (school children preferentially contact one another over adults [23], or people with many sex-partners may preferentially select partners who also have many partners); there may be multiple modes of transmission (*e.g.*, IV needle sharing, heterosexual contacts, and homosexual contacts have different contact structures and different transmission probabilities [24]); the disease may have multiple phases of infection, each with different transmission rates (*e.g.*, the acute phase of HIV is widely believed to be more infectious than the longer-lasting chronic phase [25]); or even, the partnership dynamics may evolve in response to knowledge about infection status (*e.g.*, “serosorting” in response to HIV). These and many other features have been previously studied through mathematical models.

In this paper our goals are to show that many of the effects that have been studied by other approaches under the assumption that partnerships have zero duration can be captured by the EBCM approach to incorporate the effect of long-lived partnerships and to provide guidance to other researchers working with similar questions. The resulting equations are simpler than those found by other methods, do not require approximations inherent in many competing approaches, and allow us to account for heterogeneity in numbers of partners. The flexibility of this approach allows us to develop exact models for a wide range of different situations. Our scope is to show that for each effect considered the method results in a simple system of equations which accurately reproduces simulated epidemics rather than to perform a detailed investigation of the impact of each effect. We begin by demonstrating the model in the simplest static network with minimal structure. We next provide derivations for the cases described above, summarized in table 1, which by themselves form a “recipe book” for use by other modelers. The list is by no means exhaustive, and it is not difficult to combine different effects. More generally these may be used as exemplar models which can guide the derivation of models for other behaviors not considered here. We focus primarily on static networks and assume a very small initial condition; the generalization to dynamic networks is straightforward following [1], and the generalization to allow larger initial conditions is also straightforward following [26]. For our final example of serosorting we must turn to a dynamic network model as it is fundamental to the process. We finally give a discussion of limitations of the EBCM method, giving examples for which there appears to be no simple model. We make no claim that we are the first to study these effects, most of which have been studied in a mass-action setting already. Rather, this is the first simple approach allowing us to write down exact models for these and similar effects in a network setting.

Method and Results

2.1 The basic model

The static network epidemic models we present here are all generalizations of the basic Configuration Model (CM) network epidemic model. To set the stage, we first define a CM network [27–30]. Because we will use the same underlying approach, we also briefly describe the underlying method of [1]. Generalizing the models we develop to dynamic networks is straightforward following [1].

In a CM network, each individual is represented by a *node* which is connected to other nodes by *edges* which can transmit disease. To construct the network, each node is assigned a number of edges (its *degree*) k with probability $P(k)$. The edges connect randomly to one another using proportional mixing, so that the probability of

Table 1. Models investigated in this article.

Model	Brief Description	Section
Directed Networks	Model for a disease in which some partnerships are not symmetric in terms of disease risk.	2.2.1
Heterogeneous Individuals	Model for populations with heterogeneities in infectiousness and/or susceptibility that do not correlate with population structure	2.2.2
Assortative mixing by type	Model for populations with demographic groups that have heterogeneities in infectiousness and/or susceptibility and partner selection is affected by an individual's group.	2.2.3
Multiple transmission modes	Model for a disease that can be transmitted by more than one type of behavior and the network structure induced by each behavior is different.	2.2.4
Multiple infectious stages	Model for a disease which has several infectious stages of possibly varying duration or infectiousness.	2.2.5
Serosorting	Model for dynamic network where edges break or are created at rates dependent on the status of partners.	2.3.1, 2.3.2

The edge-based compartmental models considered here. All of these except serosorting are presented using the (static) Configuration Model network structure. Serosorting is presented in two different dynamic network contexts.
doi:10.1371/journal.pone.0069162.t001

selecting a partner of degree k is $kP(k)/\langle K \rangle$ where $\langle K \rangle$ denotes the average of k . The generation of a CM network is illustrated in figure 1. It is convenient to define

$$\psi(x) = \sum_k P(k)x^k.$$

Note that $\psi'(x) = \sum_k kP(k)x^{k-1}$ and $\psi'(1) = \langle K \rangle$.

We define $S(t)$ to be the proportion of the population still susceptible at time t , $I(t)$ to be the proportion infected at time t , and $R(t)$ the proportion recovered at time t . We make the assumption that these change deterministically at the population scale. By assuming that the population scale dynamics are deterministic, we are making one key assumption that we highlight here. When we consider a single individual u in multiple realizations of the epidemic, the time of infection of u or even whether u is ever infected is a random event. However, if the disease dynamics are deterministic at the population scale, then the details of when u is infected and who u might later infect must not matter at the population scale. This is analogous to the ‘‘price-taker’’ assumption of economics in which, for example, a small farmer could not produce enough wheat to affect the price of

wheat, and so the farmer must take whatever price the market is offering. The value of the price-taker assumption mathematically is that it allows us to ignore the impact an individual firm has on the broader market and just focus on the impact of the market on the firm's actions. In our case, the corresponding assumption allows us to focus our attention on the impact of the epidemic on u and explicitly ignore the impact of u on the epidemic. Because u cannot affect the dynamics of the epidemic, we can modify u by not allowing it to transmit onwards without affecting the dynamics. When we prevent u from transmitting onwards, it has no impact on the status of u , but it simplifies the calculation of the status of u since we can then ignore correlations of partners of u that arise from transmission through u .

Consider a random *test node* u chosen uniformly from the population. We alter u so that if infected, u does not transmit to its partners. Because u is chosen randomly, the probability that it is susceptible, infected, or recovered at a given time must match S , I , and R . Thus our calculation proceeds by calculating the probability u is in each state. We focus our attention on finding the probability u is susceptible. Once we know it, we are able to determine I and R by $I = 1 - S - R$, $\dot{R} = \gamma I$.

We use θ to be the probability that a random partner v of u has not transmitted infection to u . We break θ into three parts: ϕ_S , ϕ_I , and ϕ_R , which are respectively the probability v is still susceptible, the probability v is infected but has not transmitted to u , and the probability v is recovered and did not transmit to u . We have $\theta = \phi_S + \phi_I + \phi_R$.

The flow diagram of figure 2 gives the fluxes between the various compartments. Because the infection rate within a partnership is β and the recovery rate of an individual is γ it is relatively straightforward to see that the fluxes from ϕ_I to $1 - \theta$ and ϕ_R are $\beta\phi_I$ and $\gamma\phi_I$ respectively. The calculation of the flux from ϕ_S to ϕ_I is less obvious. We have two options. We can calculate the flux directly or we can simply calculate ϕ_S explicitly as was done in [1], in which case we can avoid the flux calculation altogether. This second option is simpler, and we use it here.

We have

$$\dot{\theta} = -\beta\phi_I. \tag{1}$$

The remainder of our derivation focuses on finding ϕ_I . Because $\theta = \phi_S + \phi_I + \phi_R$, we have $\phi_I = \theta - \phi_S - \phi_R$. Thus we simply need to calculate ϕ_S and ϕ_R in terms of θ to find ϕ_I in terms of θ . To calculate ϕ_S , we use the fact that the probability a partner v has

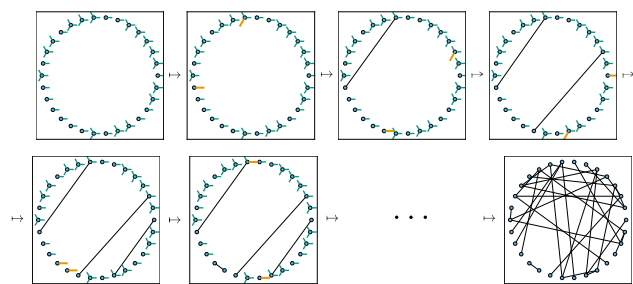


Figure 1. Sample generation of a Configuration Model network. The steps to generate a Configuration Model network with $P(1) = P(3) = 0.5$. (Top left) Each node/individual is independently assigned either 1 or 3 stubs with equal probability. (Successive plots) Pairs of stubs are randomly chosen and joined into edges until no stubs remain (bottom right). In the limit of a large population, about half of the individuals have 3 partners, about half have just 1 partner, and a random partner of a given individual is three times as likely to have 3 partners versus just 1. The number of triangles in the network is small; there is little clustering. For this $P(k)$ we have $\psi(x) = (x + x^3)/2$.
doi:10.1371/journal.pone.0069162.g001

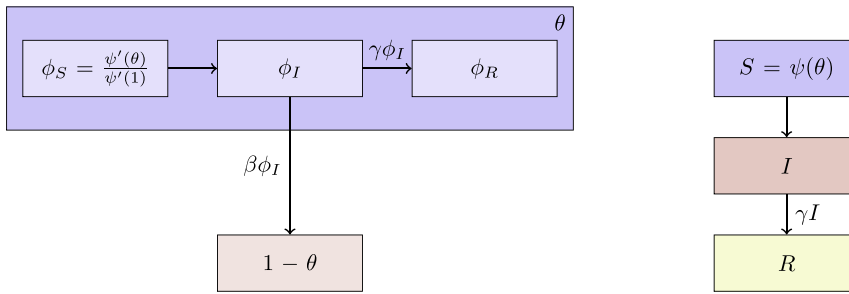


Figure 2. Flow diagram for Configuration Model networks. Given a test individual u and a random partner v , the probability v has not transmitted to u is θ . We divide θ into three compartments based on the status of v : $\theta = \phi_S + \phi_I + \phi_R$ where the subscript gives the status of v . The flow diagram on the left gives the flux between these subcompartments within θ and the flux from θ to $1 - \theta$ (which comes specifically from the ϕ_I subcompartment of θ). The flow diagram on the right shows the flux of individuals through the $S, I,$ and R compartments. To find θ , we must find ϕ_I , which is $\theta - \phi_S - \phi_R$. We find $\phi_S = \psi'(\theta)/\psi'(1)$. Because the flux into ϕ_R and $1 - \theta$ are proportional, we can find $\phi_R = \gamma(1 - \theta)/\beta$. Thus $\phi_I = \theta - \psi'(\theta)/\psi'(1) - \gamma(1 - \theta)/\beta$, and we are able to find a differential equation for θ in terms of θ . To find $S, I,$ and R , we note that the probability u is in each state is equal to the proportion of the population in each state, so the susceptible proportion is equal to the probability that u is susceptible, which is $\psi(\theta)$. We determine I by $I = 1 - S - R$ and R by $\dot{R} = \gamma I$. doi:10.1371/journal.pone.0069162.g002

degree k is $kP(k)/\langle K \rangle$. Since u is prevented from infecting v , the probability v is susceptible given its k is θ^{k-1} . Thus ϕ_S is a weighted average of this, $\phi_S = \sum_k kP(k)\theta^{k-1}/\langle K \rangle = \psi'(\theta)/\psi'(1)$.

To calculate ϕ_R , we look at figure 2. The fluxes into ϕ_R and into $1 - \theta$ are proportional to one another, with the proportionality coefficient γ/β . Since both variables begin at 0, this means $\phi_R = \gamma(1 - \theta)/\beta$. Consequently

$$\phi_I = \theta - \frac{\psi'(\theta)}{\psi'(1)} - \frac{\gamma}{\beta}(1 - \theta).$$

This leads to

$$\dot{\theta} = -\beta\theta + \beta \frac{\psi'(\theta)}{\psi'(1)} + \gamma(1 - \theta).$$

Thus we have the system of equations

$$\dot{\theta} = -\beta\theta + \beta \frac{\psi'(\theta)}{\psi'(1)} + \gamma(1 - \theta) \tag{2}$$

$$\dot{R} = \gamma I \quad S = \psi(\theta) \quad I = 1 - S - R, \tag{3}$$

which has just two ordinary differential equations (ODEs). In fact equation (2) does not depend on equation (3), so the system is governed by the single ODE (2).

We show the variables and their definitions in table 2.

2.1.1 Initial Condition. Our derivation has implicitly assumed that the epidemic is started by an infinitesimally small initial condition, so that as $t \rightarrow -\infty$, we find $\theta \rightarrow 1$, $\phi_S \rightarrow 1$, $\phi_I \rightarrow 0$, and $\phi_R \rightarrow 0$. By focusing on linear terms, we find that the vector of $1 - \phi_S, 1 - \theta, \phi_I,$ and ϕ_R can be expressed in terms of an eigenvector of the associated matrix times e^{rt} for some r . Thus if we know any one of these at $t = 0$, we can find the other variables by assuming that it has converged to this eigenvector. More generally, if the initial condition is not small, we cannot assume that it converges to the eigenvector before nonlinear terms become important. In this case, we need to account for the initial condition explicitly. This is discussed in [26]. The derivations below all assume small initial condition and that the epidemic is above the critical threshold, but the approach of [26] can account for a

larger initial condition in each case or the behavior if the disease is not able to invade.

2.1.2 Dynamic Networks. We can adapt the approach above to dynamic networks with relatively little difficulty. The main addition is that we have additional fluxes from and to the various ϕ compartments as edges break or are formed. We must include some additional variables showing the proportion of all edges that belong to nodes of a given status as this determines the probability that a new edge is with a susceptible, infected or recovered node. Details are in [1,31].

2.1.3 Generalizing the model. In the remainder of this section, we generalize the model for many static network situations, and in section 2.3 we discuss some generalizations specific to dynamic networks. The basic approach is to consider a random test node u which is prevented from causing infection. Then consider the edges which could transmit infection to u . We determine the probability the edges have not transmitted to u , which may depend on u , the partner, or details of the partnership. Our approach to determining this probability is the same as above. Once we know the probability any given edge has not transmitted to the test node, we can calculate the probability that the test node is susceptible, from which we can calculate the proportion of the population that is susceptible, infected, or recovered.

Our approach is fairly general and will apply to a wide range of populations. The generalizations we have here are by no means exhaustive, and many of the effects we consider here could be combined. The main property we require of the population is that the partners of a given individual are independent of one another. More precisely we require that we be able to assume that if we know the population scale details of the epidemic, then knowing the status of one partner of the test individual u gives no information about the status of another partner of u .

2.2 Epidemics on generalized static networks

2.2.1 Directed Networks. There are a number of realistic scenarios where infection can transmit in only one direction. Examples include blood transfusions, a food handler infecting a consumer, and even a patient infecting a doctor where they come into contact only because of the patient's infection. The probability and final size of epidemics for this scenario have been studied previously [32], but not the dynamics. Other researchers have investigated the dynamic spread of disease through asymmetric networks using pair-approximation based techniques

Table 2. Variables and equations for the basic EBCM approach.

Variable	Definition	Equation
θ	The probability that v has not transmitted to u .	$\dot{\theta} = -\beta\phi_I$
ϕ_S	The probability that v has not transmitted to u and is susceptible.	$\phi_S = \frac{\psi(\theta)}{\psi(1)}$
ϕ_I	The probability that v has not transmitted to u and is infected.	$\phi_I = \theta - \phi_S - \phi_R$
ϕ_R	The probability that v has not transmitted to u and has recovered.	$\phi_R = \frac{\gamma(1-\theta)}{\beta}$
S	The probability u is susceptible or equivalently the proportion of the population that is susceptible.	$S = \psi(\theta) = \sum P(k)\theta^k$
I	The probability u is infected or equivalently the proportion of the population that is infected.	$I = 1 - S - R$
R	The probability u is recovered or equivalently the proportion of the population that is recovered.	$\dot{R} = \gamma I$
$P(k)$	The proportion of individuals within the population having degree k .	Not Applicable

Equations and variables assuming a negligibly small initial proportion infected. In each case u is a randomly chosen test individual (prevented from transmitting to others) and v is a random partner of u .
doi:10.1371/journal.pone.0069162.t002

[33]. The resulting pair-approximation model relies on a large number of equations. The model we derive here predicts the dynamics of an epidemic in a directed network, requires only a handful of equations, and is exact for the assumed network class.

We can investigate the dynamics in almost the same manner as before. Assume that the network consists of both directed and nondirected edges. The disease can transmit along a directed edge only following the edge direction, while nondirected edges may be followed in either direction. Let β_d and β_n denote the rate of transmission along directed and nondirected edges respectively. Recovery occurs at rate γ regardless of how infection was received.

We refer to edges pointing to a node of interest as *in-directed*, and those pointing away as *out-directed* edges. The probability of having k_i in-, k_o out-, and k_n nondirected edges is given by $P(k_i, k_o, k_n)$. We define

$$\psi(x, y, z) = \sum_{k_i, k_o, k_n} P(k_i, k_o, k_n) x^{k_i} y^{k_o} z^{k_n}.$$

We again consider a random test node u which is prevented from causing infection. We define $\theta_d(t)$ and $\theta_n(t)$ to be the probability an in-directed edge or nondirected edge to u respectively has not transmitted infection to u . The probability that u is still susceptible is $\psi(\theta_d, 1, \theta_n)$. We use the variables $\phi_{S,d}$, $\phi_{I,d}$, and $\phi_{R,d}$ to be the equivalent of ϕ_S , ϕ_I , and ϕ_R seen before for in-directed edges. Following the same approach as before we arrive at the flow diagrams in figure 3.

Consider a partner v with a directed edge to u . Because of how v is chosen, the probability it has k_i in-, k_o out-, and k_n nondirected edges is $k_o P(k_i, k_o, k_n) / \langle K_o \rangle$. The probability that v is still susceptible is $\phi_{S,d} = \sum_{k_i, k_o, k_n} k_o P(k_i, k_o, k_n) \theta_d^{k_i} \theta_n^{k_o} / \langle K_o \rangle = \frac{\partial}{\partial y} \psi(\theta_d, 1, \theta_n) / \frac{\partial}{\partial y} \psi(1, 1, 1)$. The probability that v has recovered without infecting u is $\phi_{R,d} = \gamma(1 - \theta_d) / \beta_d$. Because $\phi_{I,d} = \theta_d - \phi_{S,d} - \phi_{R,d}$ we have $\phi_{I,d}$ in terms of θ_d and θ_n . This gives us $\dot{\theta}_d$ in terms of θ_d and θ_n . A similar expression holds for $\dot{\theta}_n$. We have

$$\dot{\theta}_d = -\beta_d \theta_d + \beta_d \frac{\frac{\partial}{\partial y} \psi(\theta_d, 1, \theta_n)}{\frac{\partial}{\partial y} \psi(1, 1, 1)} + \gamma(1 - \theta_d) \quad (4)$$

$$\dot{\theta}_n = -\beta_n \theta_n + \beta_n \frac{\frac{\partial}{\partial z} \psi(\theta_d, 1, \theta_n)}{\frac{\partial}{\partial z} \psi(1, 1, 1)} + \gamma(1 - \theta_n). \quad (5)$$

To this we add

$$\dot{R} = \gamma I \quad S = \psi(\theta_d, 1, \theta_n) \quad I = 1 - S - R \quad (6)$$

to give us the proportion of susceptible, infected, and recovered individuals.

If the system only has directed edges, then we can drop θ_n from the analysis and $\psi(x, y, z)$ reduces to $\psi(x, y)$. Such a model could be used to study the impact of superspreaders where the probability of receiving infection from an infected node is similar for all nodes (in-degrees are similar), but some nodes have many more partners to infect than others (high variance in out-degree). With minor modifications, we can adapt this method to edges which may transmit in both directions but have asymmetric transmission rates.

We now demonstrate the equations in a concrete example. Consider a population for which the average in-degree, out-degree, and nondirected degrees are each \hat{k} as follows: Each node has in-degree \hat{k} . The out-degree k_o is uniformly chosen from 0 up to $2\hat{k}$ inclusive, and the nondirected degree is $2\hat{k} - k_o$. For this population,

$$\psi(x, y, z) = x^{\hat{k}} \frac{z^{2\hat{k}+1} - y^{2\hat{k}+1}}{(2\hat{k} + 1)(z - y)}.$$

Figure 4 shows results for $\hat{k} = 4$, $\beta_d = 0.2$, and $\beta_n = 0.4$.

2.2.2 Heterogeneous infectiousness and susceptibility. Assume now that there is a parameter j which measures a node's ability to become infected and cause infection, but does not influence the contact structure of the population. We refer to the value of j for a node as its *type*, and the probability a node has a given type j is $Q(j)$. Although we assume the type is discretely distributed, it could be continuous with no significant complications. The recovery rate γ_j of a type- j node and the transmission rate $\beta_{j,l}$ from a type- l node to a type- j node are type dependent. The final size for a special case of this model where a node's infectiousness and susceptibility are uncorrelated is given in

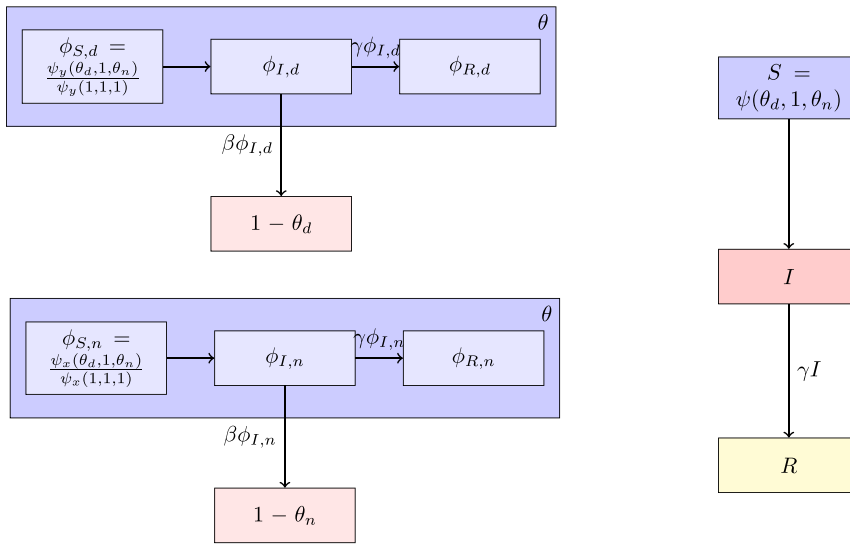


Figure 3. Directed CM model. Flow diagram for a network with directed and nondirected edges. We consider the two edge types separately. The evolution of edges is similar to figure 2. We can assign different infection rates for each edge type. doi:10.1371/journal.pone.0069162.g003

[34]. Similar problems have been considered for well-mixed mass action populations (e.g., [35,36] and many others).

Consider a random test node u of type j . Let $\theta_{j,l}$ denote the probability that an edge from a type- l partner v to u has not transmitted infection from v to u , and similarly $\phi_{S;j,l}$ the probability v is still susceptible, $\phi_{I;j,l}$ the probability v is infected but the edge has not transmitted, and $\phi_{R;j,l}$ the probability that v has recovered without transmitting. We define $\bar{\theta}_j = \sum_l \theta_{j,l} Q(l)$ as the probability that a random edge to u has not transmitted infection to u . We use the original definition of $\psi(x) = \sum_k P(k)x^k$.

We find that v is susceptible with probability $\phi_{S;j,l} = \sum_k k P(k) \bar{\theta}_l^{k-1} / \langle K \rangle = \psi'(\bar{\theta}_l) / \psi'(1)$. We also find that $\phi_{R;j,l} = \gamma_l (1 - \theta_{j,l}) / \beta_{j,l}$. Then the flow diagram in figure 5 shows that

$$\dot{\theta}_{j,l} = -\beta_{j,l} \theta_{j,l} + \beta_{j,l} \frac{\psi'(\bar{\theta}_l)}{\psi'(1)} + \gamma_l (1 - \theta_{j,l}). \quad (7)$$

The probabilities a type- j node is still susceptible, infected, or

recovered satisfy

$$\dot{R}_j = \gamma_j I_j \quad S_j = \psi(\bar{\theta}_j) \quad I_j = 1 - S_j - R_j. \quad (8)$$

The total population in each state is given by

$$S = \sum_j S_j Q(j) \quad I = \sum_j I_j Q(j) \quad R = \sum_j R_j Q(j). \quad (9)$$

We now demonstrate the equations in a concrete example. One application of this model is to the impact of a partially effective vaccination. Vaccination generally reduces the susceptibility of a node, but could either increase or decrease the infectiousness of a node by reducing the severity of symptoms (less sick individuals may shed less virus but also maintain stronger contact intensity while symptomatic). If only part of the population is vaccinated, then the population can be divided into those who have or have not received vaccination.

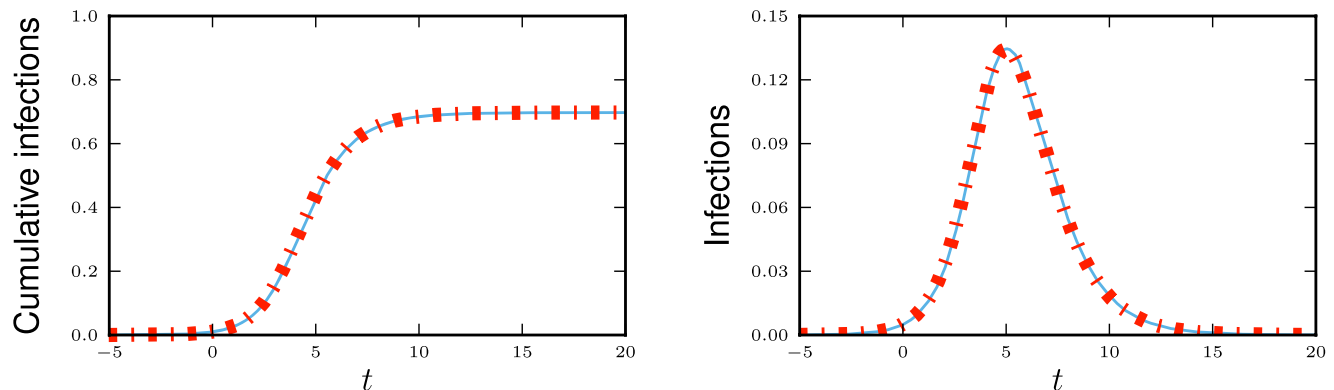


Figure 4. Directed CM example. Results for the directed networks described in section 2.2.1 using $\hat{k} = 4$, $\beta_d = 0.2$, and $\beta_n = 0.4$. We choose $t = 0$ to correspond to 1% cumulative incidence. Theory (dashed) correctly predicts dynamics of simulations in a population of 5×10^5 individuals (solid). doi:10.1371/journal.pone.0069162.g004

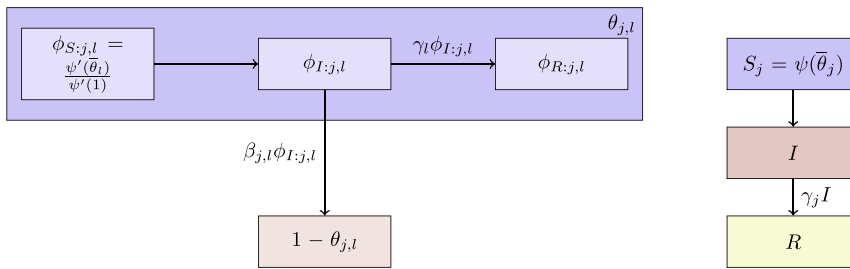


Figure 5. Heterogeneous infectiousness/susceptibility model. We separate nodes by type j , but assume that j has no effect on connectivity. Both infectiousness and susceptibility may depend on j . We must consider edges between each pair of types j and l separately. The evolution of edges is similar to before.
doi:10.1371/journal.pone.0069162.g005

Consider a population with a negative binomial degree distribution $NB(r, p)$ with size $r=3/2$ and probability $p=8/9$, giving an average degree of $pr/(1-p)=12$ and variance of $pr/(1-p)^2=108$. For a negative binomial distribution $NB(r, p)$ we have $\psi(x)=[(1-p)/(1-px)]^r$, so

$$\psi(x) = (9 - 8x)^{-3/2}.$$

Assume that half of the population has received a leaky vaccine such that vaccinated nodes have reduced susceptibility, and — if infected — reduced infectiousness and infection duration. Let γ be the rate of recovery for unvaccinated nodes and β the rate of infection between unvaccinated nodes. Vaccinated nodes recover at rate 2γ , and the rate of infection between a vaccinated and unvaccinated node (in either direction) is $\beta/2$ while the rate of infection between two vaccinated nodes is $\beta/4$. The vaccine is distributed uniformly. Results for $\beta=0.3$, $\gamma=0.5$ are shown in figure 6.

2.2.3 Populations with assortative mixing by type. In many instances there is biased mixing between or within demographic groups, and the transmission/recovery parameters for the different groups may differ. For example, the spread of influenza is strongly affected by the increased level of mixing and increased infection rates between children. Many sexually transmitted diseases are strongly affected by differences in mixing rates and risk behavior among MSM and heterosexual groups. For this reason it is useful to have a model accommodating different levels of mixing within and between groups. Others have applied a

pair-approximation approach for which assortative mixing is dependent on degree [10]. Here we generalize to also allow for assortative mixing that results from other demographic features such as age. The model we derive is equivalent (though simpler) to that of [24].

Assume that the population is made up of M groups, and let $P_j(k_1, k_2, \dots, k_M)$ denote the probability a node of group j has k_l partnerships with nodes of group l for $l=1, \dots, M$. To simplify notation, we denote this by $P_j(\mathbf{k})$ where $\mathbf{k}=(k_1, k_2, \dots, k_M)$. We similarly set $\mathbf{x}=(x_1, x_2, \dots, x_M)$ and use $\mathbf{x}^{\mathbf{k}}$ to denote $x_1^{k_1} x_2^{k_2} \dots x_M^{k_M}$. We set

$$\psi_j(\mathbf{x}) = \sum_{\mathbf{k}} P_j(\mathbf{k}) \mathbf{x}^{\mathbf{k}},$$

and let $\beta_{j,l}$ be the rate of transmission across an edge from group l to group j . We similarly define γ_j to be the recovery rate of a group j node.

We define $\theta_{j,l}$ to be the probability an edge to a test node in group j coming from a group l node has not transmitted infection. If our test node u is of type j , then the probability that a partner v of type l is still susceptible is $\phi_{S;j,l} = \sum_{\mathbf{k}} k_j P_l(\mathbf{k}) \theta_{l,1}^{k_1} \theta_{l,2}^{k_2} \dots \theta_{l,j}^{k_j-1} \dots \theta_{l,M}^{k_M} / \sum_{\mathbf{k}} k_j P_l(\mathbf{k}) = \frac{\partial}{\partial x_j} \psi_l(\boldsymbol{\theta}_l) / \frac{\partial}{\partial x_j} \psi_l(1)$ where $\boldsymbol{\theta}_l$ denotes the vector $(\theta_{l,1}, \theta_{l,2}, \dots, \theta_{l,M})$ and 1 denotes the vector $(1, 1, \dots, 1)$. We can also show that v has recovered without transmitting to u with probability $\phi_{R;j,l} = \gamma_l(1 - \theta_{j,l}) / \beta_{j,l}$.

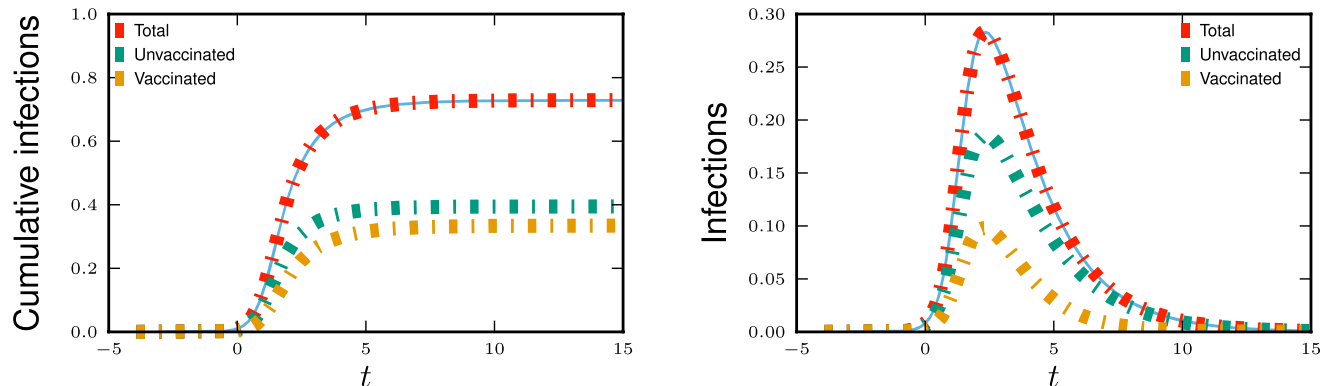


Figure 6. Heterogeneous infectiousness/susceptibility example. Epidemics spreading in a population for which half have received a leaky vaccine described in section 2.2.2. Vaccinated individuals are half as infectious and half as susceptible. We choose $t=0$ to correspond to 1% cumulative incidence. Simulations in a population of 5×10^5 individuals (solid) and theory (dashed) are in good agreement.
doi:10.1371/journal.pone.0069162.g006

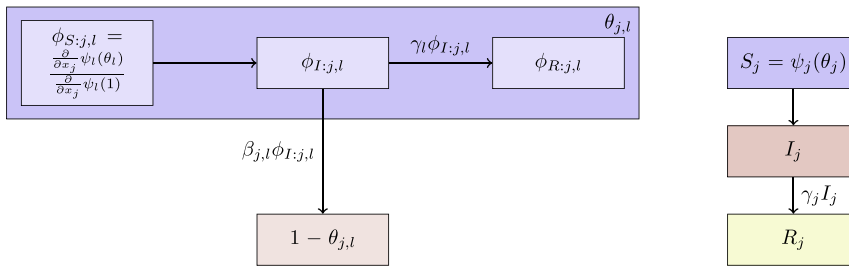


Figure 7. Assortative mixing by type model. We separate nodes by type. We assume that type may influence infectiousness and susceptibility as well as connections. For simplicity, we assume a finite number of groups. The resulting system is similar to our system for correlated infectiousness and susceptibility.
doi:10.1371/journal.pone.0069162.g007

Figure 7 gives

$$\dot{\theta}_{j,l} = -\beta_{j,l}\theta_{j,l} + \beta_{j,l} \frac{\frac{\partial}{\partial x_j} \psi_l(\theta_l)}{\frac{\partial}{\partial x_j} \psi_l(1)} + \gamma_l(1 - \theta_{j,l}). \quad (10)$$

The denominator $\frac{\partial}{\partial x_j} \psi_l(1)$ is simply the average of k_j for nodes of group l . We find

$$\dot{R}_j = \gamma_j I_j \quad S_j(t) = \psi_j(\theta_j) \quad I_j = 1 - S_j - R_j. \quad (11)$$

As a special case, we can consider a population where the number of partnerships a node has with one group is assigned independently of the number that node has with any other group. We set $P_{j,n}(k)$ to be the probability a node of group j has k edges to a node of group n and define $\psi_{j,n}(x) = \sum_k P_{j,n}(k)x^k$. Then $\psi_j(\theta_j)$ factors and may be written as $\psi_{j,1}(\theta_{j,1})\psi_{j,2}(\theta_{j,2}) \dots \psi_{j,M}(\theta_{j,M})$. In this special case we get

$$\dot{\theta}_{j,l} = -\beta_{j,l}\theta_{j,l} + \beta_{j,l} \frac{\psi'_{l,j}(\theta_{l,j})}{\psi'_{l,j}(1)} \left(\prod_{i \neq j} \psi_{l,i}(\theta_{l,i}) \right) + \gamma_l(1 - \theta_{j,l}). \quad (12)$$

We now demonstrate the equations in a concrete example. To demonstrate the ability to capture demographic information, we consider a population made up of two groups, which we arbitrarily label *children* and *adults*. The between-group degrees are binomially distributed with $\text{Bi}(4, 1/2)$, so that the average between-group degree is 2. An adult's within-group degree equals its between-group degree. In contrast, a child's within-group degree is given by 5 times its between-group degree. Thus people with higher between-group degree have higher within-group degree. We get

$$\psi_a(x_a, x_c) = (0.5x_a x_c + 0.5)^4$$

$$\psi_c(x_a, x_c) = (0.5x_a x_c^5 + 0.5)^4.$$

We set the disease parameters to be

$$\gamma_c = 1 \quad \gamma_a = 0.1$$

$$\beta_{c,c} = 0.3 \quad \beta_{c,a} = 0.1 \quad \beta_{a,a} = 0.1 \quad \beta_{a,c} = 0.1.$$

The results are shown in the top of figure 8.

We repeat with the same parameters, but this time with the correlations in degree switched so that higher between-group degrees implies lower within-group degrees. An adult's within-group degree is $4 - \hat{k}$ where \hat{k} is its between-group degree. A child's within-group degree is $5(4 - \hat{k})$ where \hat{k} is its between-group degree. We have

$$\psi_a(x_a, x_c) = (0.5x_a + 0.5x_c)^4$$

$$\psi_c(x_a, x_c) = (0.5x_a + 0.5x_c^5)^4.$$

The results are shown in the bottom of figure 8.

The distribution of within and between-group partnerships in the two populations are the same. The only distinction is that the correlations of within and between-group partnerships are different. This results in differences in the course of the epidemics. A mass action model could not distinguish between the populations.

In both cases the disease spreads quickly through the child population. Early on the spread in adults is driven largely by the explosive growth in children. Because of the correlations of adults' within and between-group degrees, those adults who are infected by children in the first scenario tend to have more adult partners and infect high-degree adults who in turn infect more high-degree children. In the second scenario however, infected children tend to infect fewer adults who tend to have fewer adult partners. The disease does not grow as quickly, but it also decays less quickly because more high degree nodes remain.

2.2.4 Multiple modes of transmission. Rather than having different types of nodes, there may be multiple modes of transmission with different mixing and infection rates for each mode. For example, HIV may spread through sexual contact and needle-sharing. The sexual contact network may have little to no relation to the needle-sharing network. Previous attempts to analyze such effects in a network context have typically assumed that there are two types of contacts: mass action-like contacts and partnerships within a network [37]. The model we derive here allows each mode of transmission to have its own underlying network structure and the equations are simpler. Assume there are M types of partnerships and that the rate of transmission along partnerships of type j is given by β_j . Let the joint distribution of the number of each partnership type be given by $P(\mathbf{k})$ where

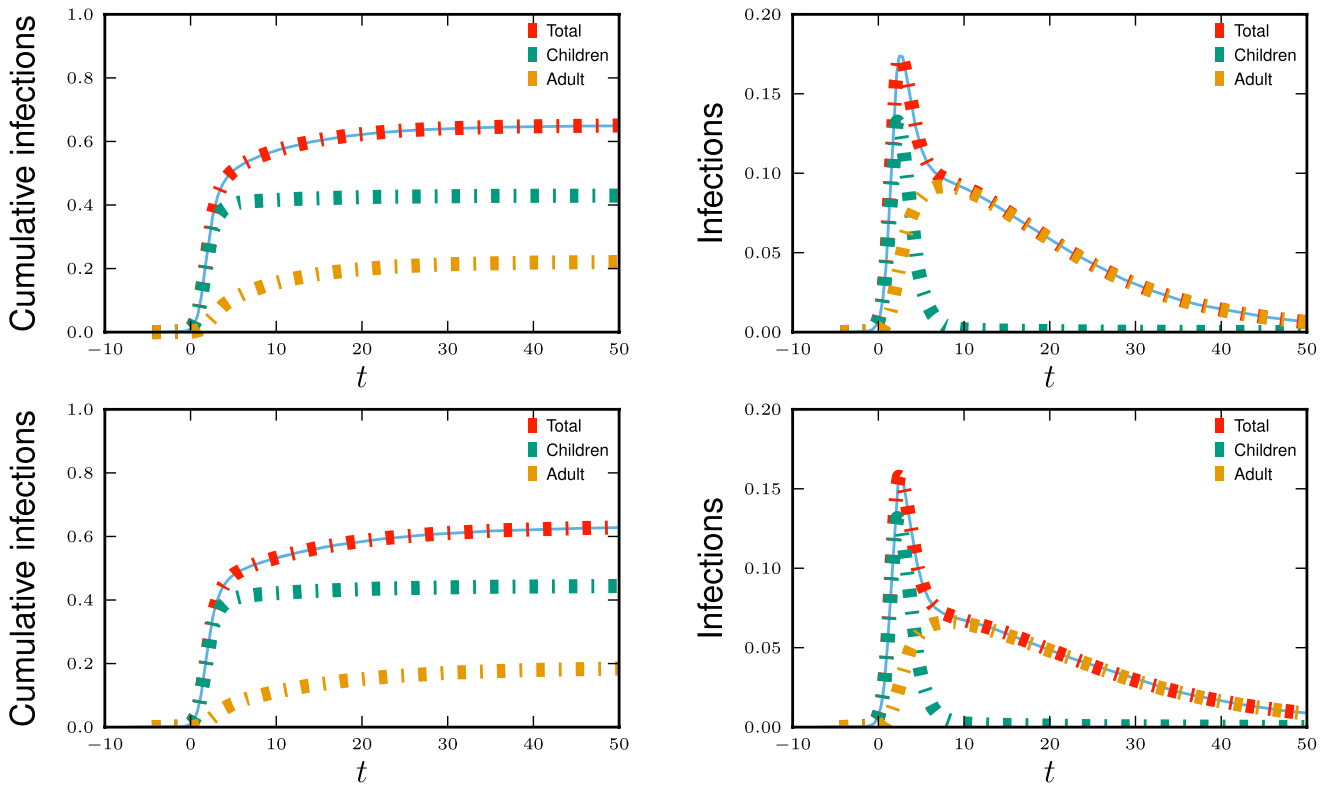


Figure 8. Assortative mixing by type example. Comparison of theory and simulated results for mixing with demographic groups described in section 2.2.3. We also show the predicted levels of infection in each subgroup. Simulations in a population of 5×10^5 individuals (solid) and theory (dashed) are in good agreement. The difference between the top and bottom result from changing the correlations of within and between-group mixing. We choose $t=0$ to correspond to 1% cumulative incidence. doi:10.1371/journal.pone.0069162.g008

$\mathbf{k}=(k_1,k_2,\dots,k_M)$. Assume recovery rates are independent of how infection was acquired. We set $\mathbf{x}=(x_1,\dots,x_M)$ and denote the product $x_1^{k_1} \dots x_M^{k_M}$ by $\mathbf{x}^{\mathbf{k}}$. We define

$$\psi(\mathbf{x}) = \sum_{\mathbf{k}} P(\mathbf{k}) \mathbf{x}^{\mathbf{k}}.$$

We can apply the same method to each edge type as shown in figure 9. We set $\phi_{S,j}$ to be the probability that an edge of type j connects the test node to a susceptible partner. If θ_j is the probability a partnership of type j has not yet transmitted infection, we set $\boldsymbol{\theta}=(\theta_1,\theta_2,\dots,\theta_M)$. We find $\phi_{S,j} = \frac{\partial}{\partial x_j} \psi(\boldsymbol{\theta}) / \frac{\partial}{\partial x_j} \psi(\mathbf{1})$

$\frac{\partial}{\partial x_j} \psi(\mathbf{1})$ and similarly the probability an edge of type j connects to a recovered partner which did not transmit is $\phi_{R,j} = \gamma(1-\theta_j) / \beta_j$. As before we find $S(t) = \psi(\boldsymbol{\theta})$. We have

$$\dot{\theta}_j = -\beta_j \theta_j + \beta_j \frac{\frac{\partial}{\partial x_j} \psi(\boldsymbol{\theta})}{\frac{\partial}{\partial x_j} \psi(\mathbf{1})} + \gamma(1-\theta_j) \quad (13)$$

$$\dot{R} = \gamma I \quad S = \psi(\boldsymbol{\theta}) \quad I = 1 - S - R. \quad (14)$$

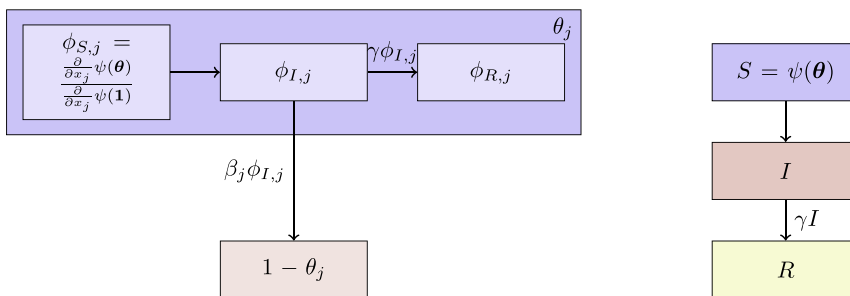


Figure 9. Multiple modes of transmission model. Flow diagram showing the flux of edges for the j -th contact type for a disease which has multiple modes of transmission. doi:10.1371/journal.pone.0069162.g009

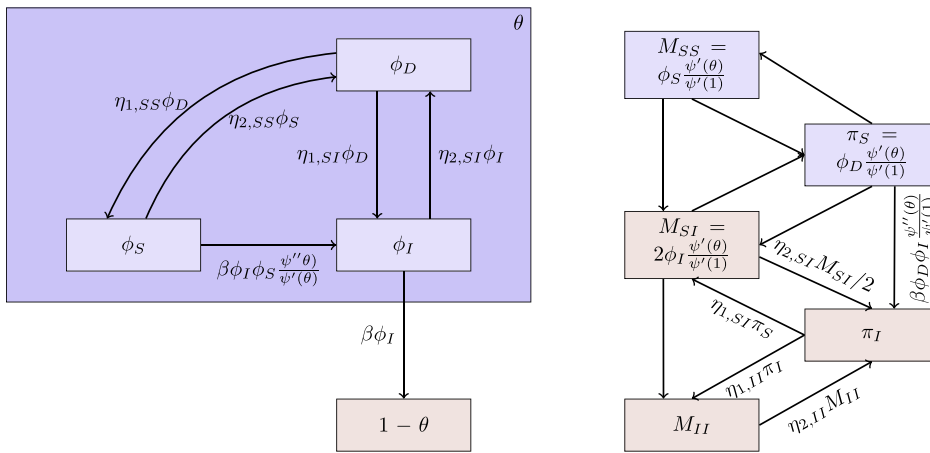


Figure 13. Fixed-degree serosorting model. Flow diagram showing the interplay involved in serosorting with fixed-degree. We do not consider a recovered class, which simplifies the equations. The framework can be adapted to include a recovered class. The M variables represent the total proportion of stubs involved in edges between the two types and the π variables are the proportion of dormant stubs belonging to nodes of each type. The ϕ variables are as before. For the right hand side, we are able to determine most of the variables analytically, so we only need the fluxes into and out of π_I . We expect that the edge breaking and rejoining rates η will depend on values of π_S and π_I . doi:10.1371/journal.pone.0069162.g013

In the case where the degree with respect to one partnership type is independent of that with respect to another, we can simplify equation (13) to be similar to equation (12).

We now demonstrate the equations in a concrete example. We consider a population with three different types of partnerships. We take k_1, k_2 , and k_3 to denote the number of each type of partnership a node has. We assume that partnership type 1 is binomially distributed with $\text{Bi}(2, 0.5)$ (giving a mean of 1). Partnership type 2 is geometrically distributed, with mean 2. Partnership type 3 has a negative binomial distribution $\text{NB}(1, 1/4)$ with mean $1/3$ and variance $4/9$. The numbers of partners an individual has of each type are assigned independently. We find

$$\psi(x_1, x_2, x_3) = \frac{(x_1 + 1)^2}{4} \frac{x_2}{2 - x_2} \frac{3}{4 - x_3}.$$

We take $\gamma = 1$ and set $\beta_1 = 1, \beta_2 = 0.5$, and $\beta_3 = 3$ for each partnership type. We compare simulation and theory in figure 10. This example is chosen so that the variables of ψ can be separated into different terms. The theory still applies even when this factorization of ψ is not possible.

2.2.5 Multiple infectious stages. There are a number of diseases with multiple infectious stages such as Tuberculosis and HIV. Some diseases begin with a non-infectious latent phase, some begin with a highly infectious acute stage before settling into a long-term chronic stage, and others oscillate between phases of high and low infectiousness. To model such situations we adapt a standard chain progression model, for which there are M infectious phases shown in figure 11. We are not able to explicitly solve for all variables in terms of θ , so we must find the fluxes between the compartments. We can still find $\phi_S = \psi'(\theta)/\psi'(1)$, so we are able to find $\phi_{I,1}$ in terms of the other variables using

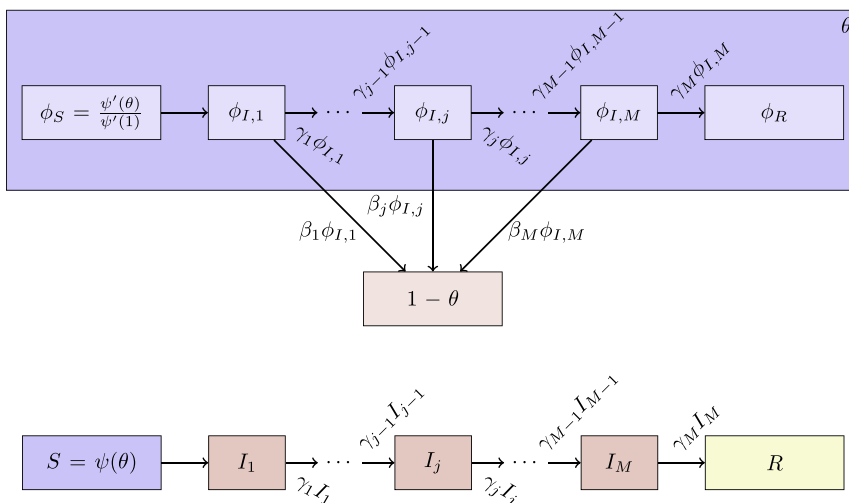


Figure 11. Multiple infectious stages model. Flow diagrams for a disease with several infected stages. When a disease progresses through several states (or has an infectious period that is not exponentially distributed) it is convenient to use a stage-progression model to represent the state of an edge. doi:10.1371/journal.pone.0069162.g011

$\phi_S + \phi_R + \sum_j \phi_{I_j} = \theta$. We obtain

$$\dot{\theta} = - \sum \beta_j \phi_{I_j} \tag{15}$$

$$\phi_{I,1} = \theta - \frac{\psi'(\theta)}{\psi'(1)} - \phi_R - \sum_{j=2}^M \phi_{I_j} \tag{16}$$

$$\dot{\phi}_{I_j} = \gamma_{j-1} \phi_{I_{j-1}} - \gamma_j \phi_{I_j} - \beta_j \phi_{I_j} \quad M \geq j > 1 \tag{17}$$

$$\dot{\phi}_R = \gamma_M \phi_{I,M} \tag{18}$$

$$\dot{I}_j = \gamma_{j-1} I_{j-1} - \gamma_j I_j \quad M \geq j > 1 \tag{19}$$

$$\dot{R} = \gamma_M I_M \quad S = \psi(\theta) \quad I_1 = 1 - S - R - \sum_{j=2}^M I_j, \tag{20}$$

where I_j is the proportion of the population in the j -th infectious class.

We now demonstrate the equations in a concrete example. Consider now the spread of a disease for which there are three infectious stages. The first stage is moderately infectious but not long, with $\beta_1 = 0.2$, $\gamma_1 = 1$. The second stage is much longer, but has a substantially lower infectiousness, with $\beta_2 = 0.01$ and $\gamma_2 = 0.08$. The final stage has an intermediate duration but substantially higher infectiousness, with $\beta_3 = 2$ and $\gamma_3 = 0.4$. We assume that the disease spreads in a population with degree distribution $NB_{(1,4/5)}$ having mean 4 and variance 20 with

$$\psi(x) = (5 - 4x)^{-1}.$$

The results are shown in figure 12.

2.3 Dynamic Networks and Serosorting

For some diseases, it is not uncommon for individuals to actively seek out partners of similar disease status (as in HIV [38] or

Leprosy [39]) or even discordant status (as in “chicken pox parties” or “swine flu parties”). This is commonly known as serosorting. To study these populations, we must use dynamic network models, which we developed in [1].

We study serosorting in two models. In the first, we use an “actual degree” model where an individual has a given number of potential partnerships, of which only a proportion are active at any given time. In the second, we use an “expected degree” model in which individuals break any existing partnerships at a fixed rate, but different individuals may find new partners at differing rates leading to a variation in the expected number of partnerships across the population.

For simplicity, we will assume that there is no recovered class and once infected an individual remains infected. This restriction is easily removed, but by using it, we are able to simplify the model and reduce the number of parameters needed. We again consider a test node, and as before we assume that if infected the test node does not cause any infections. We make an additional assumption that if infected the test node continues to behave as if it were susceptible, and that its potential partners treat it as if it were susceptible. We can think of the test node as an individual who is immune, but is unaware of that immunity, and we track the probability that the test node has not yet received a sufficient dose to infect a non-immune individual.

2.3.1 Actual Degree Serosorting model. In the actual degree formulation, we think of an individual as having k stubs or half-edges. These stubs may be active (and connected to another node’s stub) or dormant (and available to form new edges).

When an edge breaks, the corresponding stubs enter a dormant phase. We assume that the rate a dormant edge belonging to a susceptible individual finds a new susceptible partner is $\eta_{1,SS}$, and the rate it finds a new infected partner is $\eta_{1,SI}$. These may depend on the density of susceptible and infected individuals in the population. Similarly, active edges break at rates depending on the status of the nodes. Edges between susceptible nodes break at rate $\eta_{2,SS}$, edges between a susceptible and infected node breaks at rate $\eta_{2,SI}$, and edges between infected nodes break at rate $\eta_{2,II}$.

We define ϕ_S , ϕ_I , ϕ_R , and ϕ_D to be the probability that a stub belonging to the test node u has never been part of an edge that transmitted infection to u , and that the stub is currently connected to a susceptible, infected, or recovered node or is dormant respectively. The fluxes between these states are shown in figure 13. Unlike in previous cases, we are unable to explicitly calculate ϕ_S , so we must track the flux into and out of ϕ_S . The fluxes between

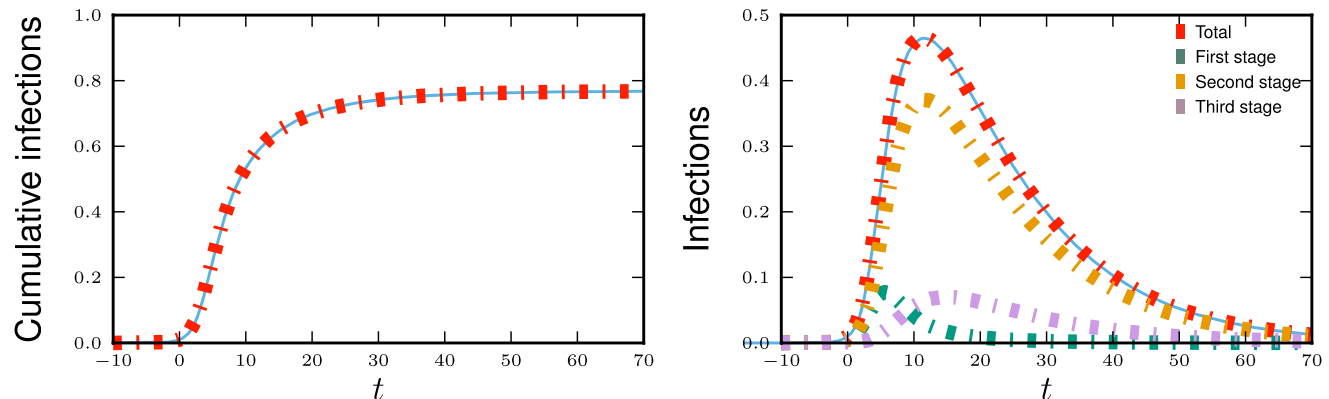


Figure 12. Multiple infectious stages example. The spread of the disease described in 2.2.5 with three infectious stages. Simulations in a population of 5×10^5 individuals (solid) and theory (dashed) are in good agreement. We choose $t=0$ to correspond to 1% cumulative incidence. doi:10.1371/journal.pone.0069162.g012

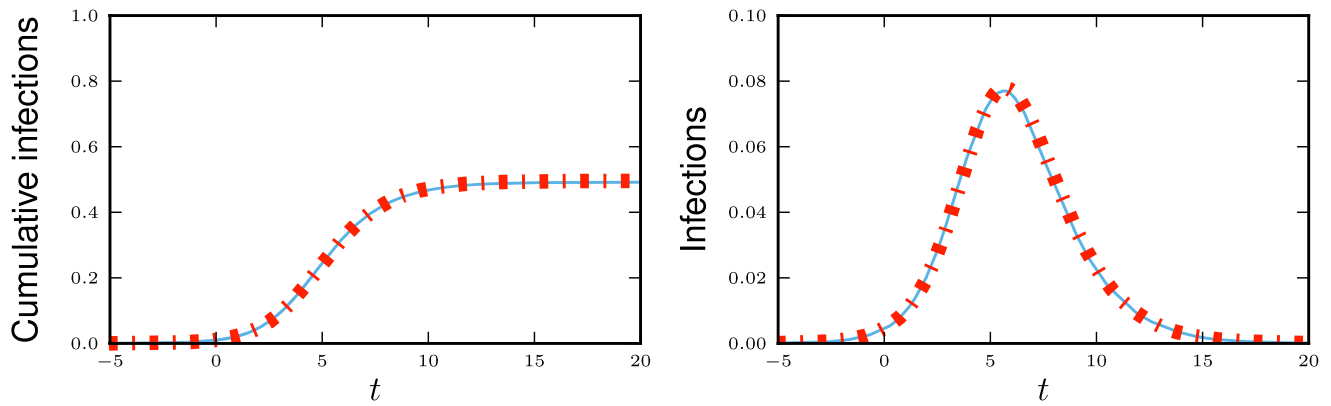


Figure 10. Multiple modes of transmission example. Disease spread in a population with three different types of partnerships, each with a different degree distribution described in 2.2.4. Simulations in a population of 5×10^5 individuals (solid) and theory (dashed) are in good agreement. We choose $t=0$ to correspond to 1% cumulative incidence. doi:10.1371/journal.pone.0069162.g010

ϕ_S and ϕ_D are straightforward. However, the flux from ϕ_S to ϕ_I requires more attention. We repeat our derivation from [1]. Consider a partner v of the test node u having the following properties: the stub belonging to u never transmitted to u prior to joining with the stub belonging to v , and similarly the stub belonging to v never transmitted to v prior to joining with the stub belonging to u . Given this, the probability v is susceptible is $q = \sum_k kP(k)\theta^{k-1} / \langle K \rangle = \psi'(\theta) / \psi'(1)$. Thus, given that v is susceptible, the rate v becomes infected is

$$-\frac{\dot{q}}{q} = -\frac{\dot{\theta}\psi'(\theta)/\psi'(1)}{\psi'(\theta)/\psi'(1)} = \beta\phi_I \frac{\psi''(\theta)}{\psi'(\theta)}.$$

Thus the flux from ϕ_S to ϕ_I is $\beta\phi_S\phi_I\psi''(\theta)/\psi'(\theta)$, the product of ϕ_S , the probability that a stub has not transmitted infection to the test node and connects to a susceptible node, with $\beta\phi_I\psi''(\theta)/\psi'(\theta)$, the rate that the partner becomes infected given that the stub has not transmitted and connects to a susceptible node. We need to account for the number of stubs that are in edges between different classes of nodes or are dormant. We use M_{SS} to be the proportion of all stubs that are in edges between susceptible nodes. Equivalently this is the probability that a stub is active, connects to a susceptible node, and belongs to a susceptible node: $M_{SS} = \phi_S\psi'(\theta)/\psi'(1)$. We similarly define M_{SI} to be the proportion of all stubs that are in edges between a susceptible and infected node. We calculate this by finding the probability a stub is active, connects to an infected node, but belongs to a susceptible node but we must multiply by 2 because this only counts one stub in each edge. We get $M_{SI} = 2\phi_I\psi'(\theta)/\psi'(1)$. We also define M_{II} to be the proportion of all stubs in edges between infected nodes. We will calculate its value later. We set π_S to be the proportion of stubs that are dormant and belong to susceptible nodes and π_I to be the proportion of stubs that are dormant and belong to infected nodes. The value of π_S can be calculated as for the dormant contact case of [1] to be $\pi_S = \phi_D\psi'(\theta)/\psi'(1)$. The value of π_I is calculated by finding the fluxes out of the other states. Once we have all of these variables, we have $M_{II} = 1 - M_{SS} - M_{SI} - \pi_S - \pi_I$.

Following figure 13 we find

$$\dot{\theta} = -\beta\phi_I \quad (21)$$

$$\dot{\phi}_S = \eta_{1,SS}(\pi_S, \pi_I)\phi_D - \eta_{2,SS}(\pi_S, \pi_I)\phi_S - \beta\phi_I\phi_S \frac{\psi''(\theta)}{\psi'(\theta)} \quad (22)$$

$$\dot{\phi}_I = \eta_{1,SI}(\pi_S, \pi_I)\phi_D - \eta_{2,SI}(\pi_S, \pi_I)\phi_I - \beta\phi_I + \beta\phi_I\phi_S \frac{\psi''(\theta)}{\psi'(\theta)} \quad (23)$$

$$\dot{\phi}_D = \eta_{2,SS}(\pi_S, \pi_I)\phi_S + \eta_{2,SI}(\pi_S, \pi_I)\phi_I - [\eta_{1,SS}(\pi_S, \pi_I) + \eta_{1,SI}(\pi_S, \pi_I)]\phi_D \quad (24)$$

$$\pi_S = \phi_D \frac{\psi'(\theta)}{\psi'(1)} \quad (25)$$

$$\dot{\pi}_I = \frac{\eta_{2,SI}(\pi_S, \pi_I)}{2} M_{SI} - \pi_S \eta_{1,SI}(\pi_S, \pi_I) + \eta_{2,II}(\pi_S, \pi_I) M_{II} - \pi_I \eta_{1,II} + \beta\phi_D \phi_I \frac{\psi''(\theta)}{\psi'(1)} \quad (26)$$

$$M_{SS} = \phi_S \frac{\psi'(\theta)}{\psi'(1)} \quad (27)$$

$$M_{SI} = 2\phi_I \frac{\psi'(\theta)}{\psi'(1)} \quad (28)$$

$$M_{II} = 1 - M_{SS} - M_{SI} - \pi_S - \pi_I, \quad (29)$$

where $\eta_{1,SS}$, $\eta_{1,SI}$, $\eta_{1,II}$, $\eta_{2,SS}$, $\eta_{2,SI}$, and $\eta_{2,II}$ are likely to depend on π_I and π_S and may depend on the other quantities. What form that dependence takes is determined by the behavior of the population. This completes the derivation of the equations for the actual degree formulation of serosorting.

2.3.2 Variable Degree Serosorting model. In many populations, it is reasonable to assume that individuals create

and break contacts without regard to whether contacts already exist. Consequently, the concept of having a fixed number of stubs is inappropriate. For these populations, we assume that in the absence of disease all contacts will have the same expected duration but different individuals will create new contacts at different rates, resulting in some having higher or lower average degrees. In [1], we used κ to be the *expected degree* of a node. However, when behavior changes based on infection status, the expected degree of individuals can change. Instead, we refer to κ as the *desired* degree because depending on how sorting happens, it may not be possible for a node to have expected degree κ . However, κ will represent the expected degree of an individual if there were no infection present. We again use S and I to be proportions of the population. Π_S and Π_I measure the proportion of desired contacts which belong to susceptible or infected individuals: $\Pi_S = \int s_\kappa \kappa \rho(\kappa) d\kappa$, $\Pi_I = \int i_\kappa \kappa \rho(\kappa) d\kappa$ where s_κ and i_κ denote the proportion of individuals with desired degree κ who are susceptible and infected respectively.

We assume that the population behavior proceeds as before, but an uninfected node will end and form contacts with different rates for infected or susceptible partners. There are many ways in which this could be modeled. We will assume that a susceptible individual with desired degree κ acquires new susceptible contacts at rate $\kappa \eta_{1,S} \Pi_S$ and new infected contacts at rate $\kappa \eta_{1,I} \Pi_I$, where both η parameters may depend on Π_S and Π_I . Similarly, a susceptible individual will end an existing contact with another susceptible and with an infected individual at rates $\eta_{2,S}$ and $\eta_{2,I}$ respectively (where again both η parameters may depend on Π_S and Π_I). We assume that $\eta_{1,S}$ and $\eta_{2,S}$ are equal if $\Pi_I = 0$ so that in a disease-free population an individual's expected and desired degrees coincide.

We need to add variables in order to track the probability of having existing edges connecting to susceptible or infected nodes. Consider a test node u_1 with desired degree κ , and another u_2 with desired degree $\kappa + \Delta\kappa$. We define $\Phi_S \Delta\kappa$ to be the expected additional number of edges to susceptible partners that u_2 would have and $\Phi_I \Delta\kappa$ to be the expected additional number of edges to infected partners which have not transmitted that u_2 would have. We take the values of Φ_S and Φ_I in the $\Delta\kappa \rightarrow 0$ limit. In the cases considered in [1], the value of Φ_S and Π_S were the same. However, because there is active selection of partner based on disease status, in this case $\Phi_S \neq \Pi_S$.

The resulting flow diagram is shown in figure 14. We must find the flux from Φ_S to Φ_I . Consider a random test node u and look at a randomly chosen susceptible partner v . Given the desired degree κ_v of v , the rate that v becomes infected is $\beta \Phi_I \kappa_v$. We need to determine the expected value of κ_v given that v is a susceptible partner of u . We first note that the probability density function for the partner to be susceptible and have degree κ is proportional to $q(\kappa) = e^{-\kappa(1-\Theta)} \kappa \rho(\kappa) / \langle K \rangle$ with some proportionality constant a . So in order to calculate the expected value of the desired degree we take $\int a \kappa q(\kappa) d\kappa / \int a q(\kappa) d\kappa$. This simplifies to $\Psi''(\Theta) / \Psi'(\Theta)$. So the flux from Φ_S to Φ_I due to infection of the partner is $\beta \Phi_S \Phi_I \Psi''(\Theta) / \Psi'(\Theta)$

We find

$$\dot{\Theta} = -\beta \Phi_I \tag{30}$$

$$\dot{\Phi}_S = \eta_{1,S}(\Pi_S, \Pi_I) \Pi_S - \eta_{2,S}(\Pi_S, \Pi_I) \Phi_S - \beta \Phi_S \Phi_I \frac{\Psi''(\Theta)}{\Psi'(\Theta)} \tag{31}$$

$$\dot{\Phi}_I = \eta_{1,I}(\Pi_S, \Pi_I) \Pi_I - [\eta_{2,I}(\Pi_S, \Pi_I) + \beta] \Phi_I + \beta \Phi_S \Phi_I \frac{\Psi''(\Theta)}{\Psi'(\Theta)} \tag{32}$$

$$\Pi_S = \frac{\Psi'(\Theta)}{\Psi'(1)} \quad \Pi_I = 1 - \Pi_S \tag{33}$$

$$S = \Psi(\Theta) \quad I = 1 - S. \tag{34}$$

So an SI epidemic in a population with serosorting can be captured by a system with just three ODEs.

Discussion

We have applied the edge-based compartmental model approach introduced in [1] to diseases and populations with different structures, showing how to derive models for a range of scenarios. We have shown that the models derived accurately reproduces simulated epidemics in large populations with CM-like

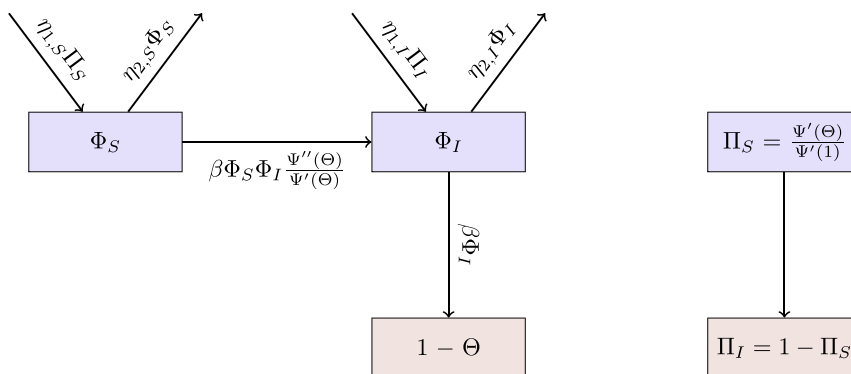


Figure 14. Variable-degree serosorting model. Flow diagram showing the interplay involved in serosorting. We do not consider a recovered class, which simplifies the equations significantly. The framework can be adapted to include a recovered class. The Π variables give the proportion of contacts that would be formed with susceptible or infected individuals assuming that their behavior is not altered by disease. The Φ variables are the probability that a current contact of the test node is with an individual of given type, under the assumption that the test node always behaves as if susceptible, and does not transmit to its partners. We expect that the edge breaking and rejoining rates η will depend on values of Π_S and Π_I . Note that $\Phi_S + \Phi_I$ need not equal Θ . doi:10.1371/journal.pone.0069162.g014

structure. With the exception of serosorting we focused our attention on static CM networks. We have considered each variation in isolation. However it is possible to adapt the approach to a disease for which several of these issues are considered simultaneously in any of the network classes discussed in [1].

In general, we can adapt most existing mean-field/mass-action style SIR models in a closed population to the spread of infectious disease through a network. When we do this, we get a ϕ variable corresponding to each of the S , I , or R variables in the usual model. We take the usual flow diagram for S , I , and R and adapt

it to give the fluxes between the ϕ variables. We add one more compartment $1-\theta$, and flux goes from each of the potentially infectious ϕ variables to $1-\theta$. This approach produces an accurate model for disease spread through the modeled population.

Author Contributions

Conceived and designed the experiments: JCM EMV. Performed the experiments: JCM EMV. Analyzed the data: JCM. Wrote the paper: JCM.

References

- Miller JC, Slim AC, Volz EM (2012) Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface* 9: 890–906.
- Miller JC (2011) A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology* 62: 349–358.
- Volz EM (2008) SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology* 56: 293–310.
- Decreusefond L, Dhersin JS, Moyal P, Tran VC (2012) Large graph limit for an SIR process in random network with heterogeneous connectivity. *The Annals of Applied Probability* 22: 541–575.
- Anderson RM, May RM (1991) *Infectious Diseases of Humans*. Oxford: Oxford University Press.
- May RM, Anderson RM (1988) The transmission dynamics of human immunodeficiency virus (HIV). *Phil Trans R Soc Lond B* 321: 565–607.
- May RM, Lloyd AL (2001) Infection dynamics on scale-free networks. *Physical Review E* 64: 066112.
- Moreno Y, Pastor-Satorras R, Vespignani A (2002) Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 26: 521–529.
- Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Physical Review Letters* 86: 3200–3203.
- Eames K, Keeling M (2002) Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences* 99: 13330–13335.
- Eames K, Keeling M (2006) Coexistence and specialization of pathogen strains on contact networks. *The American Naturalist* 168: 230–241.
- Bauch C (2002) A versatile ode approximation to a network model for the spread of sexually transmitted diseases. *Journal of mathematical biology* 45: 375–395.
- House T, Davies G, Danon L, Keeling MJ (2009) A motif-based approach to network epidemics. *Bulletin of Mathematical Biology* 71: 1693–1706.
- House T, Keeling M (2011) Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface* 8: 67–73.
- Miller JC, Kiss IZ (2013) Epidemic spread in networks: Existing methods and current challenges. Under Review.
- Ball F, Neal P (2008) Network epidemic models with two levels of mixing. *Mathematical Biosciences* 212: 69–87.
- Lindquist J, Ma J, van den Driessche P, Willeboordse F (2011) Effective degree network disease models. *Journal of Mathematical Biology* 62: 143–164.
- Diekmann O, De Jong MCM, Metz JAJ (1998) A deterministic epidemic model taking account of repeated contacts between the same individuals. *Journal of Applied Probability* 35: 448–462.
- Altmann M (1998) The deterministic limit of infectious disease models with dynamic partners. *Mathematical Biosciences* 150: 153–175.
- Dietz K, Tudor D (1992) Triangles in heterosexual hiv transmission. *AIDS epidemiology: methodological issues* : 143–155.
- Dietz K, Hadelar K (1988) Epidemiological models for sexually transmitted diseases. *Journal of Mathematical Biology* 26: 1–25.
- Lloyd-Smith JO, Getz WM, Westerhoff HV (2004) Frequency-dependent incidence in models of sexually transmitted diseases: portrayal of pair-based transmission and effects of illness on contact behaviour. *Proceedings of the Royal Society of London Series B: Biological Sciences* 271: 625–634.
- Wallinga J, Teunis P, Kretzschmar M (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology* 164: 936.
- Volz E, Frost S, Rothenberg R, Meyers L (2010) Epidemiological bridging by injection drug use drives an early HIV epidemic. *Epidemics* 2: 155–164.
- Pilcher C, Tien H, Eron J, Vernazza P, Leu S, et al. (2004) Brief but efficient: acute HIV infection and the sexual transmission of HIV. *Journal of Infectious Diseases* 189: 1785.
- Miller JC Epidemics on networks with large initial conditions or changing structure. Under Review.
- Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms* 6: 161–179.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Review* 45: 167–256.
- Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Physical Review E* 64: 026118.
- van der Hofstad R (2010) *Random Graphs and Complex Networks*. URL <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>. Author's website: accessed 18 Jun 2013.
- Miller JC, Volz EM (2012) Model hierarchies in edge-based compartmental modeling for infectious disease spread. *Journal of Mathematical Biology* : 1–31.
- Meyers LA, Newman M, Pourbohloul B (2006) Predicting epidemics on directed contact networks. *Journal of Theoretical Biology* 240: 400–418.
- Sharkey K, Fernandez C, Morgan K, Peeler E, Thrush M, et al. (2006) Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *Journal of Mathematical Biology* 53: 61–85.
- Miller JC (2007) Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E* 76: 010101(R).
- Ball F, Clancy D (1995) The final outcome of an epidemic model with several different types of infective in a large population. *Journal of Applied Probability* : 579–590.
- Ball F (1985) Deterministic and stochastic epidemics with several kinds of susceptibles. *Advances in Applied Probability* 17: 1–22.
- Ball F, Neal P (2002) A general model for stochastic SIR epidemics with two levels of mixing. *Mathematical Biosciences* 180: 73–102.
- Parsons J, Schrimshaw E, Wolitski R, Halkitis P, Purcell D, et al. (2005) Sexual harm reduction practices of HIV-seropositive gay and bisexual men: serosorting, strategic positioning, and withdrawal before ejaculation. *AIDS* 19: S13.
- Leviticus, chapter 13.