SIR dynamics in random networks with heterogeneous connectivity

Erik Volz

Abstract Random networks with specified degree distributions have been proposed as realistic models of population structure, yet the problem of dynamically modeling SIR-type epidemics in random networks remains complex. I resolve this dilemma by showing how the SIR dynamics can be modeled with a system of three nonlinear ODE's. The method makes use of the probability generating function (PGF) formalism for representing the degree distribution of a random network and makes use of network-centric quantities such as the number of edges in a well-defined category rather than node-centric quantities such as the number of infecteds or susceptibles. The PGF provides a simple means of translating between network and node-centric variables and determining the epidemic incidence at any time. The theory also provides a simple means of tracking the evolution of the degree distribution among susceptibles or infecteds. The equations are used to demonstrate the dramatic effects that the degree distribution plays on the final size of an epidemic as well as the speed with which it spreads through the population. Power law degree distributions are observed to generate an almost immediate expansion phase yet have a smaller final size compared to homogeneous degree distributions such as the Poisson. The equations are compared to stochastic simulations, which show good agreement with the theory. Finally, the dynamic equations provide an alternative way of determining the epidemic threshold where large-scale epidemics are expected to occur, and below which epidemic behavior is limited to finite-sized outbreaks.

Keywords Epidemic disease · SIR · Networks · Degree distribution

Mathematics Subject Classification (2000) 92D30 · 92D25

E. Volz (✉)
Department of Integrative Biology, University of Texas, Austin, TX, USA
e-mail: erik.volz@mail.utexas.edu

1 Introduction

Contact patterns constitute an important aspect of heterogeneity within a population of susceptible and infectious individuals, but it has been a difficult factor to incorporate into epidemiological models. Compartment models can capture many aspects of population heterogeneity, such as with respect to heterogeneous susceptibility and infectiousness [1,9,32], however such models usually assume individuals mix homogeneously within each category. In contrast, the contact patterns responsible for the spread of many infectious diseases tend to be characterized by constant relationships marked by high levels of heterogeneity in the number of contacts per individual.

An alternative approach is to model a population of susceptibles and infecteds and the contact patterns among them as a static random network [13,18,26,31]. This approach has generated a new category of epidemiological models in which epidemics spread from node to node by traversing network connections [2,8,17,19,24,25, 29,30,33]. Random networks with specified degree distributions have been proposed as a simple but realistic models of population structure. This case has the advantage of being well understood mathematically. The expected final size of epidemics in random networks with a given degree distribution has been solved exactly [19,24]. The network approach has the advantage that the mathematics of stochastic branching processes [4,15,34] can be brought to bear on the problem. This allows for precise descriptions of the mean outbreak size well as the final size.

A shortcoming of the network model is that it is difficult to describe the explicit dynamical behavior of epidemics on networks. The mean outbreak size is easy to calculate, yet the dynamic epidemic incidence, that is the number of new infecteds at a time t, has been difficult to derive. Simulation has been used in this case [12].

Heterogeneity in the number of contacts within networks makes it difficult to derive differential equations to describe the course of an epidemic. Nevertheless, several researchers [5,7,27,28] have been successful modeling many of the dynamical aspects of network epidemics, particularly in the early stage where asymptotically correct equations for disease incidence are known. These solutions break down, however, when the finite size of a population becomes a significant factor. Other work has focused on pair approximation and moment closure [6,11]. These methods produce viable approximations to SIR dynamics, but are typically high-dimensional and computationally intensive.

Here we introduce an alternative strategy based on a low-dimensional system of nonlinear ordinary differential equations. This model can be used to solve for epidemic incidence at any time, from an initial infected to the final size as well as other quantities of interest. We treat the simplest possible case of the SIR dynamics with constant rate of infection and recovery. Section 2 describes the model. Several examples are given in Sect. 3, and Sect. 3.1 compares the analytical results to stochastic simulations.

2 SIR in random networks

The networks considered here are random networks with an arbitrary degree distribution p_k (p_k being the probability of a random node having degree k) [23,26]. Nodes

can be in any of three exclusive states: susceptible ($\mathcal{S}$), infectious ($\mathcal{I}$), or recovered ($\mathcal{R}$). The dynamics are as follows. When a node is infectious, it will transmit infection to each of its neighbors independently at a constant rate r. Infectious nodes become recovered at a constant rate μ, whereupon they will no longer infect any neighbors. This will be made precise in the next section.

It is desirable to determine the dynamics of the number of susceptibles and infecteds and to develop equations in terms of those quantities. This, however, turns out to be intractable due to heterogeneity in the number of contacts. The problem can be resolved by developing equations in terms of dynamic variables representing network-based quantities, for example, the number of connections to susceptible or infectious nodes at a time t. The network- and node-based quantities are defined in the next section.

To bridge the divide between connection- and node-based quantities, a mathematical device known as a probability generating function (PGF) [34] is extremely useful. The PGF has many useful properties and is frequently used in probability theory and the theory of stochastic branching processes. Given a discrete probability density p_k, the PGF is defined as the series:

$$g(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \cdots . \tag{1}$$

The variable x in the generating function serves only as a place-holder. To illustrate the utility of this device, consider the possibility that the probability of a node being infected, say λ, is compounded geometrically according the node's degree. Then, the probability of a degree k node being susceptible is $(1 - \lambda)^k$, that is, the probability of not being infected along any of k connections. If the hazard is identical for all nodes, the cumulative epidemic incidence (the fraction of nodes infectious or recovered) will be

$$J = 1 - [p_0(1 - \lambda)^0 + p_1(1 - \lambda)^1 + p_2(1 - \lambda)^2 + \cdots] \tag{2}$$
$$= 1 - g(1 - \lambda). \tag{3}$$

Table 1 gives a summary of the parameters used in the model.

2.1 Definitions

An undirected network can be defined as a graph $\mathcal{G} = \{V, \mathcal{E}\}$ consisting of a set of vertices V corresponding to the nodes in the network, and a set of edges $\mathcal{E}$ with elements of unordered pairs of vertices, $\{a, b\}$ where $a, b \in V$. Two vertices a, b are said to be *neighbors* or *neighboring each other* or simply *connected* if $\{a, b\} \in \mathcal{E}$. For the purposes of this model, the terms "vertex" and "node" will often be used interchangeably.

The networks considered here can be generated by the Configuration Model(CM) [22]. For our purposes, the important aspect of CM networks is that the probability of being connected to a node is proportional to the degree of the node. Denote the degree of a node $v \in V$ as d_v. Then given an edge $\{a, x\} \in \mathcal{E}$, the probability that

Table 1 Parameters and dynamic variables for the network SIR model

- $r :=$ Force of infection. The constant rate at which infectious nodes infect a neighbor
- $\mu :=$ Recovery rate. The constant rate at which infected nodes become recovered
- $p_k :=$ The probability that a node will have degree k
- $g(x) :=$ The probability generating function for the degree distribution $\{p_k\}$
- $S :=$ The fraction of nodes susceptible at time t
- $I :=$ The fraction of nodes infectious at time t
- $R :=$ The fraction of nodes recovered at time t
- $J = I + R$ The cumulative epidemic incidence at time t
- $\mathcal{A}_X$ Set of arcs (*ego*, *alter*) such that node *ego* is in set X
- M_X Fraction of arcs in set $\mathcal{A}_X$
- $\mathcal{A}_{XY}$ Set of arcs (*ego*, *alter*) s.t. *ego* $\in X$ and *alter* $\in Y$
- M_{XY} Fraction of arcs in set $\mathcal{A}_{XY}$

$x = b$ is $d_b / \sum_{i \in V} d_i$. Note that this allows multiple edges to the same node as well as loops from a node to itself, however the existence of multiple edges and loops is exceedingly rare for large sparse random networks such that results based on this case can be safely applied to networks without multiple edges. Networks of this type can be generated by a the following algorithm:

1. To each node $v \in V$ assign an i.i.d. degree δ_v from distribution p_k
2. Generate a new set X of "half-edges" with δ_v copies of node v for all nodes
3. Insure X has an even number of elements, for example, by deleting a uniform random element if odd.
4. While X is not empty, draw two elements v_1, v_2 uniformly at random and create edge $\{v_1, v_2\}$.

At any point in time, a vertex can be classified as susceptible, infectious, or recovered. Let $\mathcal{S}, \mathcal{I}$, and $\mathcal{R}$ denote the disjoint sets of vertices classified as susceptible, infectious, or recovered respectively. $\mathcal{J} = \mathcal{I} \cup \mathcal{R}$ will denote the set of infectious or recovered nodes. S, I, and R will denote the fraction of nodes in the sets $\mathcal{S}, \mathcal{I}$, and $\mathcal{R}$ respectively. The cumulative epidemic incidence will be the fraction of nodes in set $\mathcal{J}$.

As stated in the previous section, infectious vertices $a \in \mathcal{I}$ will infect neighboring susceptible vertices $b \in \mathcal{S}$ at a constant rate r. Infectious vertices will become recovered (move to set $\mathcal{R}$) at a constant rate μ.

Although the network is undirected in the sense that any two neighboring vertices can transmit infection to one another, we wish to keep track of who infects who. Therefore, for each edge $\{a, b\} \in \mathcal{E}$, let there be two arcs, which will be defined to be the ordered pairs (a, b) and (b, a). Let $\mathcal{A}$ denote the set of all arcs in the network. The first element in the ordered pair (a, b) will frequently be called the *ego* and the second element the *alter*.

$\mathcal{A}_{XY}$ will denote the subset of arcs such that *ego* $\in X$ and *alter* $\in Y$. $\mathcal{A}_X$ will denote the subset of arcs such that *ego* $\in X$. $M_{XY} = \#\{\mathcal{A}_{XY}\}/\#\{\mathcal{A}\}$ will denote the fraction of arcs in the corresponding set $\mathcal{A}_{XY}$.

Table 2 Network-based dynamic variables for the network SIR model

- $\theta :=$ The fraction of degree one nodes that remain susceptible at time t
- $p_I := M_{SI}/M_S$. The probability that an arc with a susceptible ego has an infectious alter
- $p_S := M_{SS}/M_S$. The probability that an arc with a susceptible ego has a susceptible alter

Table 3 A summary of the nonlinear differential equations used to the describe the spread of a simple SIR type epidemic through a random network. The degree distribution of the network is generated by $g(x)$

$$\dot{\theta} = -rp_I\theta$$

$$\dot{p}_I = rp_S p_I \theta \frac{g''(\theta)}{g'(\theta)} - rp_I(1 - p_I) - p_I\mu$$

$$\dot{p}_S = rp_S p_I \left(1 - \theta \frac{g''(\theta)}{g'(\theta)}\right)$$

$$S = g(\theta)$$

$$\dot{I} = rp_I\theta g'(\theta) - \mu I$$

For example, two variables will be especially important in the derivations that follow. M_{SS} is the fraction of arcs with a susceptible ego and a susceptible alter. M_{SI} is the fraction of arcs with a susceptible ego and and infectious alter. M_S will be the fraction of arcs with a susceptible ego and an alter of any type.

2.2 Dynamics

Our objective is to develop a deterministic model to describe epidemic dynamics expressed with a low-dimensional system of differential equations. At first, this goal may seem incompatible with network-SIR dynamics described in the previous section. Infection spreads along links in a random network, which implies the epidemic incidence at any time as well as the final size must also be random, depending on the particular structure of a given random network. This is true, however it is possible to avoid such considerations by focusing on epidemic dynamics in the limit as population size goes to infinity. This strategy has been used in previous work to calculate the expected final size of epidemics in infinite random networks [24] expressed as a fraction of the total population size. A similar strategy is followed here by considering the fraction of nodes in sets $\mathcal{S}, \mathcal{I}$, and $\mathcal{R}$, after a small fraction ϵ nodes are infected initially in a susceptible population.[1] The conclusion is the system of equations given in Table 3 in terms of the dynamic variables given in Table 2. The dynamics predicted by these equations are compared to stochastic simulations with large but finite networks in Sect. 3.1.

Consider a susceptible node *ego* at time t with a degree k. Then there will be a set of k arcs $\{(ego, alter_1), (ego, alter_2), \cdots, (ego, alter_k)\}$ corresponding to *ego*. We will assume that for each arc $(ego, alter_i)$ there will be a uniform probability $p_I = M_{SI}/M_S$ that $alter_i$ is infectious. Then there is an expected number kp_I arcs $(ego, alter)$ such that $alter$ is infectious. In a time dt, an expected number $rkp_I\, dt$ of these will be such that the infectious alter transmits to *ego*. Consequently, the hazard for ego becoming infected at time t is

---

[1] Although it has not yet been proven, there is computational evidence that a continuous-time stochastic SIR epidemic will converge to our deterministic model in the limit of large population size.

$$\lambda_k(t) = rkp_I(t). \tag{4}$$

Now let $u_k(t)$ represent the fraction of degree k nodes that remain susceptible at time t, or equivalently the probability that *ego* in the previous example is susceptible. Using Eq. 4,

$$u_k(t) = \exp\left\{-\int_0^t \lambda_k(\tau)d\tau\right\} = \exp\left\{-\int_0^t rkp_I(\tau)d\tau\right\}$$

$$= \exp\left\{-\int_0^t rp_I(\tau)d\tau\right\}^k. \tag{5}$$

Subsequently we will use the symbol θ to denote $u_1 = \exp\{-\int_0^t rp_I(\tau)d\tau\}$. From Eq. 5 it is clear that $u_k = \theta^k$.

Given θ, it is easy to determine the fraction of nodes which remain susceptible at a time t.

$$S = p_0 + p_1u_1 + p_2u_2 + p_3u_3 \cdots$$
$$= p_0 + p_1\theta + p_2\theta^2 + p_3\theta^3 + \cdots = g(\theta) \tag{6}$$

This equation makes use of the generating function $g(\cdot)$ for the degree distribution which greatly simplifies this and subsequent equations.

The dynamics of θ are dependent on the hazard λ_1.

$$\frac{d\theta/dt}{\theta} = -\lambda_1(t) \Rightarrow$$
$$\dot{\theta} = -\theta\lambda_1(t) = -\theta\, r\, p_I \tag{7}$$

Unfortunately, this does not completely specify the dynamics of θ and by extension S, which also depends on the variable p_I. The derivation of the dynamics of p_I follows.

$$\dot{p}_I = \frac{d}{dt}\frac{M_{SI}}{M_S} = \frac{\dot{M}_{SI}}{M_S} - \frac{\dot{M}_S M_{SI}}{M_S^2} \tag{8}$$

Our goal is to put Eq. 8 in terms of the variables θ, p_S, p_I and the PGF $g(\cdot)$. M_S is easily placed in terms of these variables.

$$M_S = \sum_k p_k \times k \times \Pr[\text{degree k node susceptible}] / \sum_k kp_k$$

$$= \sum_k p_k k\theta^k / g'(1) = \left[\frac{d}{dx}g(\theta x)\right]_{x=1} / g'(1) = \theta g'(\theta) / g'(1) \tag{9}$$

M_{SI} follows easily.

$$M_{SI} = M_S \times M_{SI}/M_S = M_S p_I = p_I \theta g'(\theta)/g'(1) \tag{10}$$

Next consider $\dot{M}_S$. In time dt, $-\dot{S}$ nodes become infectious. Since $S = g(\theta)$,

$$\dot{S} = \frac{d}{dt} S = \frac{d}{dt} g(\theta) = \dot{\theta} g'(\theta) = -r p_I \theta g'(\theta). \tag{11}$$

Calculating $\dot{M}_{SI}$ requires careful consideration of the rearrangement of arcs among sets $\mathcal{A}_{SS}$ and $\mathcal{A}_{SI}$ as $-\dot{S}$ nodes become infected in a small time interval. Since the hazard of becoming infected is proportional to the number of arcs to an infectious alter, a susceptible node will be selected with probability proportional to the number of arcs from the node to infectious nodes.

To clarify subsequent calculations, I will introduce the notation δ_{XY} to represent the average degree of nodes in set X, selected with probability proportional to the number of arcs to nodes in set Y, not counting one arc to nodes of type Y. For example, if we select an arc ($ego \in X$, $alter \in Y$) uniformly at random out of the set of arcs from nodes in set X to nodes in set Y ($\mathcal{A}_{XY}$), and follow it to the node in set X, (ego), then δ_{XY} will represent the average number of arcs ($ego, alter'$) not counting the arc we followed to ego. This is commonly called the "excess degree" of a node [20]. Furthermore, $\delta_{XY}(Z)$ will be as δ_{XY} but counting only arcs from ego to nodes in set Z, ($ego, alter \in Z$).

To calculate $\dot{M}_{SI}$ we need to first calculate δ_{SI}, and for this it is necessary to derive the degree distribution among susceptible nodes. Considering two arcs ($ego, alter_1$) and ($ego, alter_2$) with $ego \in \mathcal{S}$, we must suppose that the event that $alter_1 \in X$ is independant of the event that $alter_2 \in Y$.[2] A consequence of this is that arcs from a susceptible ego to nodes in sets $\mathcal{S}, \mathcal{I}, \mathcal{R}$ are distributed multinomially with probabilities p_S, p_I, and $p_R = 1 - p_S - p_I$ respectively. Let $d_{ego}(X)$ be the r.v. denoting the number of arcs from ego to nodes in set X. Letting c normalize the distribution, and letting the dummy variables x_S, x_I, and x_R correspond to the number of arcs from a susceptible ego to an alter in sets $\mathcal{S}, \mathcal{I}, \mathcal{R}$ respectively, the degree distribution for susceptible nodes will be generated by

$$g_S(x_S, x_I, x_R) = \sum_{k \geq 0} p_k u_k \sum_{i,j|i+j \leq k} x_S^i x_I^j x_R^{k-i-j} \Pr[d_{ego}(S) = i, d_{ego}(I) = j | p_S, p_I]/c. \tag{12}$$

---

[2] Although a rigorous proof for this is currently lacking, it is borne out by the success of this model in predicting epidemic final size and dynamics (see Sects. 3 and 3.1 below). Any topological feature which creates correlations between arcs would invalidate this derivation. This includes clustering (transitivity).

Using the multinomial theorem this becomes

$$g_S(x_S, x_I, x_R) = \sum_{k \geq 0} p_k \theta^k (x_S p_S + x_I p_I + x_R(1 - p_S - p_I))^k / c$$

$$= g(\theta(x_S p_S + x_I p_I + x_R(1 - p_S - p_I)))/g(\theta), \qquad (13)$$

where $c = \sum_k p_k \theta^k (p_S + p_I + (1 - p_S - p_I))^k = g(\theta)$ normalizes the distribution.

The degree distribution for susceptible nodes selected with probability proportional to the number of arcs to infectious nodes will be generated by the following equation. This distribution is usually called the *excess degree distribution*. For example, selecting a random arc in set $\mathcal{A}_{SI}$, following it to the susceptible node, and then counting the composition of arcs out of that node (not counting the one we arrived on) would give rise to this distribution. Note that this equation *does not* count one arc to infectious nodes (the one used to select the susceptible node).

$$g_{SI}(x_S, x_I, x_R)$$
$$= \frac{\sum_{k \geq 0} p_k u_k \sum_{i,j|i+j \leq k} j \times x_S^i x_I^{j-1} x_R^{k-i-j} \Pr[d_{ego}(S) = i, d_{ego}(I) = j | p_S, p_I] /}{\sum_{k \geq 0} p_k u_k \sum_{i,j|i+j \leq k} j \times \Pr[d_{ego}(S) = i, d_{ego}(I) = j | p_S, p_I]}$$

$$= \left[\frac{d}{dx_I} g_S(x_S, x_I, x_R) \right] / \left[\frac{d}{dx_I} g_S(x_S, x_I, x_R) \right]_{x_S = x_I = x_R = 1}$$

$$= g'(\theta(x_S p_S + x_I p_I + x_R(1 - p_S - p_I)))/g'(\theta) \qquad (14)$$

Because arcs are distributed multinomially to nodes in sets $\mathcal{S}, \mathcal{I}, \mathcal{R}$, we have $g_{SS}(x_S, x_I, x_R) = g_{SI}(x_S, x_I, x_R)$, which is easy to verify by repeating the calculation in Eq. 14.

A useful property of PGF's is that the mean of the distribution they generate can be calculated by differentiating and evaluating with the dummy variables set to one [34]. Now using Eqs. 13 and 14, we have the following results.

$$\delta_{SI} = \left[\frac{d}{dx} g_{SI}(x, x, x) \right]_{x=1} = \theta g''(\theta)/g'(\theta) \qquad (15)$$

$$\delta_{SI}(I) = \left[\frac{d}{dx_I} g_{SI}(x_S, x_I, x_R) \right]_{x_S = x_I = x_R = 1} = p_I \theta g''(\theta)/g'(\theta) \qquad (16)$$

$$\delta_{SI}(S) = \left[\frac{d}{dx_S} g_{SI}(x_S, x_I, x_R) \right]_{x_S = x_I = x_R = 1} = p_S \theta g''(\theta)/g'(\theta) \qquad (17)$$

As a fraction $-\dot{S}$ nodes leave set $\mathcal{S}$ in time dt, the fraction of arcs between $\mathcal{S}$ and $\mathcal{I}$, M_{SI} is reduced by the fraction of arcs from infectious nodes to the $-\dot{S}$ newly infectious nodes (recall $M_{SI} = M_{IS}$). Therefore M_{SI} is reduced at rate $-\dot{S}\delta_{SI}(I)/g'(1)$. Because $\delta_{SI}(I)$ does not count the arc along which a node was infected, M_{SI} is also reduced at a rate $r M_{SI}$ to account for all arcs which have an infectious ego which

transmits to the susceptible alter. And in time dt, μI nodes become recovered. The average number of arcs in $\mathcal{A}_{IS}$ per infectious node is proportional to M_{SI}/I. Then M_{SI} is reduced at a rate $\mu I(M_{SI}/I) = \mu M_{SI}$.

The quantity M_{SI} is also increased, as new infected nodes have links to susceptible nodes. A newly infectious node will have on average $\delta_{SI}(S)$ arcs to susceptible nodes, so M_{SI} is increased at a rate $-\dot{S}\delta_{SI}(S)/g'(1)$.

To summarize, M_{SI} decreases at the rates

– $-\dot{S}\delta_{SI}(I)/g'(1)$
– rM_{SI}
– μM_{SI}

And M_{SI} increases at the rates

– $-\dot{S}\delta_{SI}(S)/g'(1)$

Then applying Eqs. 16, 17, and 11 we have

$$\dot{M}_{SI} = ((-\dot{S})\delta_{SI}(S) - (-\dot{S})\delta_{SI}(I))/g'(1) - (r + \mu)M_{SI}$$
$$= rp_I(p_S - p_I)\theta^2 g''(\theta)/g'(1) - (r + \mu)M_{SI}. \tag{18}$$

Finally, it is necessary to determine the time derivative of M_S.

$$\dot{M}_S = \frac{d}{dt}\theta g'(\theta)/g'(1) = (\dot{\theta}g'(\theta) + \theta\dot{\theta}g''(\theta))/g'(1)$$
$$= (-rp_I\theta g'(\theta) - rp_I\theta^2 g''(\theta))/g'(1) \tag{19}$$

Now applying Eqs.[3] 9, 18, and 19 to Eq. 8 we solve for $\dot{p}_I$ in terms of the PGF and θ.

$$\dot{p}_I = rp_I p_S \theta \frac{g''(\theta)}{g'(\theta)} - p_I(1 - p_I)r - p_I\mu \tag{20}$$

This equation makes use of the variable p_S which changes in time. Deriving the dynamics of this variable will complete the model. This calculation is very similar to that for $\dot{p}_I$.

$$\dot{p}_S = \frac{d}{dt}\frac{M_{SS}}{M_S} = \frac{\dot{M}_{SS}}{M_S} - \frac{\dot{M}_S M_{SS}}{M_S^2} \tag{21}$$

The calculation for $\dot{M}_{SS}$ is very similar to that for $\dot{M}_{SI}$. Newly infected nodes have on average $\delta_{SI}(S)$ arcs to other susceptibles, so that

---

[3] The normalizing constant $g'(1)$ cancels out and could have been left out these equations.

$$\dot{M}_{SS} = -2 \times (-\dot{S})\delta_{SI}(S)/g'(1)$$
$$= -2rp_I p_S \theta^2 g''(\theta)/g'(1), \tag{22}$$

where the factor of $2\times$ accounts for two arcs per edge.

Now applying Eqs. 9, 19, and 22 to Eq. 21, we have

$$\dot{p}_S = rp_I p_S \left(1 - \theta \frac{g''(\theta)}{g'(\theta)}\right). \tag{23}$$

The complete system of equations is summarized in Table 3.

The fraction of infectious nodes can be solved by introducing a fourth dynamic variable. The infectious class increases at a rate $-\dot{S}$ (Eq. 11) and decreases at a rate μI. Therefore

$$\dot{I} = rp_I \theta g'(\theta) - \mu I. \tag{24}$$

An advantage of dynamic modeling of epidemics in networks is that the time-evolution of variables besides cumulative incidence can be calculated. Above it was shown how to calculate the degree distribution among susceptible nodes (eqn. 13). Additionally, the degree distribution among nodes which are either infectious or recovered (set $\mathcal{J}$) can be calculated by taking the complement.

$$g_J(x) = (g(x) - g(\theta x))/(1 - g(\theta)) \tag{25}$$

2.3 Initial conditions

If a small fraction ϵ of the nodes in the network are selected uniformly at random and initially infected, we can anticipate the following initial conditions. The fraction of arcs with infectious ego will also be $M_I = \epsilon$, and since ϵ is small, there is low chance of two initial infecteds being connected. Therefore $M_{SI} \approx M_I = \epsilon$. θ, which can be interpreted as the fraction of degree one nodes remaining susceptible will be $1 - \epsilon$. We'll assume that no initial infecteds are connected to eachother, so that $M_{SI} = M_I$. And $M_S = 1 - M_{SI} = 1 - \epsilon$ because there are initially no recovered nodes. Finally, $M_{SS} = M_S - M_{SI} = 1 - 2\epsilon$. To summarize,

1. $\theta(t = 0) = 1 - \epsilon$
2. $p_I(t = 0) = M_{SI}/M_S = \epsilon/(1 - \epsilon)$
3. $p_S(t = 0) = M_{SS}/M_S = (1 - 2\epsilon)/(1 - \epsilon)$

2.4 Epidemic threshold

Epidemic dynamics can fall into one of two qualitatively different regimes. Below a threshold in the ratio r/μ, the final size (I_∞) is necessarily proportional to the fraction of initial infectious nodes: $I_\infty \propto \epsilon$. But above this threshold, epidemics occur, and necessarily occupy a fraction of the population even as $\epsilon \to 0$.

As per Eq. 4, the number of new infections in a small time interval is proportional to p_I. This is in contrast to compartment models in which the number of new infections is proportional the current number of infectious. If $\dot{p}_I(t = 0) < 0$, an epidemic will necessarily die out without reaching a fraction of the population. The epidemic threshold occurs where

$$\dot{p}_I(t = 0) = 0 = r p_S p_I \theta \frac{g''(\theta)}{g'(\theta)} - r p_I (1 - p_I) - p_I \mu. \tag{26}$$

Applying the initial conditions above and considering $\epsilon \ll 1$ gives

$$\dot{p}_I(t = 0) = r \frac{1 - 2\epsilon}{1 - \epsilon} \frac{\epsilon}{1 - \epsilon} (1 - \epsilon) g''(\theta)/g'(\theta) - r \frac{\epsilon}{1 - \epsilon} \frac{1 - 2\epsilon}{1 - \epsilon} - \mu \frac{\epsilon}{1 - \epsilon}$$
$$= \epsilon \left(r \frac{g''(\theta)}{g'(\theta)} - r - \mu \right) = 0. \tag{27}$$

Rearranging yields the critical ratio r/μ in terms of the PGF.

$$(r/\mu)^* = \frac{g'(1)}{g''(1) - g'(1)} \tag{28}$$

The epidemic threshold in Eq. 28 can also be put in terms of the the transmissibility, which is the probability that an infectious ego will transmit infection to a given alter. If *ego* is infectious for a duration T, the probability of transmitting to a given alter is $1 - e^{-rT}$. Integrating over an exponentially distributed duration of infectiousness T, the transmissibility τ is calculated to be

$$\tau = E[1 - e^{-rT}]$$
$$= \int_0^\infty (1 - e^{-rT})(\mu e^{-\mu T}) dT = \frac{r}{r + \mu}. \tag{29}$$

Then rearranging Eq. 28 yeilds the epidemic threshold in terms of τ.

$$\tau^* = g'(1)/g''(1) \tag{30}$$

This is consistent with previous results based on bond-percolation theory [24].[4]

3 Examples

The model has been tested on several common degree distributions:

---

[4] Although the original derivation in [24] was slightly flawed, the threshold has subsequently been shown to be correct [10].

– Poisson: $p_k = \frac{z^k e^{-z}}{k!}$. This is generated by

$$g(x) = e^{z(x-1)}. \tag{31}$$

– Power-law. For our experiments, we utilize power-laws with exponential cutoffs κ: $p_k = \frac{k^{-\gamma} e^{-k/\kappa}}{Li_\gamma(e^{-1/\kappa})}, k \geq 1$ where $Li_n(x)$ is the nth polylogarithm of x. This is generated by

$$g(x) = Li_\gamma(xe^{-1/\kappa})/Li_\gamma(e^{-1/\kappa}). \tag{32}$$

– Discrete exponential: $p_k = (1 - e^{-1/\lambda})e^{-k/\lambda}$. This is generated by

$$g(x) = \frac{1 - e^{-1/\lambda}}{1 - xe^{-1/\lambda}}. \tag{33}$$

Figure 1 shows the cumulative incidence for each of the degree distributions (31), (32), and (33), with a force of infection $r = .2$ and recovery rate $\mu = .1$. Initially $\epsilon = 10^{-4}$ nodes are infected. The parameters of the degree distributions were chosen so that each network has an identical average degree of 3. That is, the density of connections in each network is the same. Nevertheless, there is widely different epidemic behavior due to the different degree distributions. Consistent with previous research, the degree distribution has a great impact on the final size of the epidemic [19,24]. More importantly, the three networks exhibit widely varying dynamical behavior. The power law network experiences epidemics which accelerate very rapidly. Such epidemics enter the expansion phase (the time at which cumulative incidence increases at its maximum rate) virtually as soon as the first individual in the network is infected. In contrast, the Poisson network experiences a lag before the expansion phase of the epidemic. The discrete exponential network has behavior similar to that of the power law but not as extreme; the expansion phase is early, but not instantaneous, and the final size is not as great as the power law. These observations are consistent with the findings in [5] that the timescale of epidemics shortens with increasing contact heterogeneity. This has important implications for intervention strategies, as it is often the case that interventions are planned and implemented only after a pathogen has circulated in the population for some time. If an epidemic were to occur in the power law network, there would be little time to react before the the infection had reached a large proportion of the population.

Recall from Sect. 2.4 that below the epidemic threshold τ^*, only small, finite-sized outbreaks will occur. Figure 2 shows the qualitatively different dynamical behavior of outbreaks below the phase transition for networks with a Poisson distribution. Below the phase transition, the final size is always proportional to the fraction of initial infecteds ϵ.

Something offered by this model and not to the author's knowledge seen previously, is an explicit calculation for how the degree distribution of susceptibles evolves over the course of the epidemic. We expect the degree distribution to become bottom-heavy, as high degree nodes are gradually weeded out of the population of susceptibles. This is indeed observed in Fig. 3 for the Poisson trial described above.
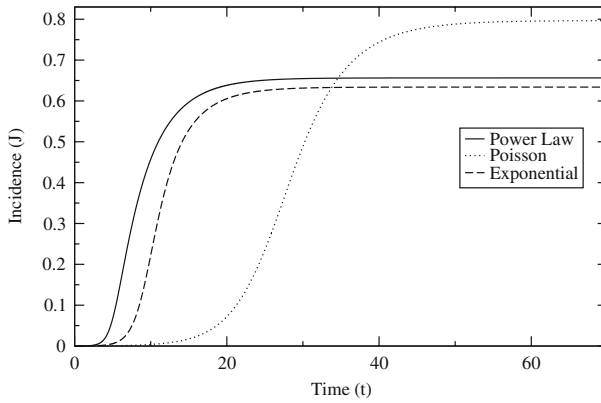
Fig. 1 The number of infecteds (including recovered) is shown versus time for an SIR model on three networks. Force of infection and mortality are constant: $r = 0.2$, $\mu = 0.1$. The networks have Poisson ($z = 3$), power law ($\gamma = 1.615$, $\kappa = 20$), and exponential ($\lambda = 3.475$) degree distributions. Each of these degree distributions has an average degree of 3

Fig. 2 The number of infecteds (including recovered) is shown versus time for an SIR model on a Poisson network ($z = 3$). Each of these trials are below the epidemic threshold required to sustain an epidemic $((r/\mu)^* = 1/2$ from Eq. 28). The outbreak size is reported as a multiple of the fraction of initial infecteds in the network. Mortality is constant, $\mu = 0.4$, while three different levels of the force of infection are tried, $r = 0.15, 0.17, 0.18$
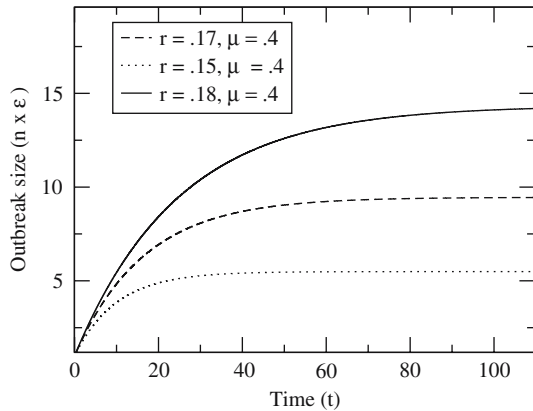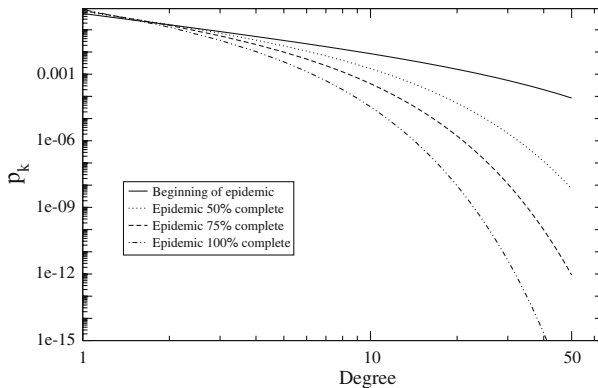

Recall that the degree distribution of susceptibles is generated by the multi-variate PGF (13). The explicit degree distribution can be retrieved from Eq. 13 by differentiation. The following gives the probability that a susceptible node has m links.

$$p_m^S = [\frac{d^k}{dx^k} g_S(x, x, x)]_{x=0} / k! \tag{34}$$

For example, applying this to the Poisson PGF (Eq. (31)) gives

$$p_k = \frac{(z\theta)^k e^{-z\theta}}{k!} \tag{35}$$

which is simply the Poisson distribution with an adjusted parameter $z \times \theta$. Another example is illustrated in Fig. 3, which shows the degree distribution among susceptibles for the power-law network considered above.

Fig. 3 The degree distribution for susceptible nodes where the epidemic size is 50, 75, and 100% of the final size, as well as degree distribution at the beginning of the epidemic. The degree distribution for the network as a whole is a power law with exponential cutoff (Eq. 32)

3.1 Stochastic simulations

Simulation of SIR on networks presents two challenges: A random network must be generated with the desired degree distribution. And, the stochastic rules that govern the transmission of disease at the microscopic scale must be well-defined.

In the deterministic model of the previous sections, we described dynamics using the quantities S, I, and R to represent the fraction of nodes in each category. The stochastic dynamics will be described by S_n, I_n, and R_n which will denote the random *number* of susceptible, infectious, and recovered nodes respectively. Below we will give evidence that the stochastic dynamics match the deterministic dynamics in terms of time-scale and final size. Future work should provide a formal characterization of this relationship. We hypothesize that stochastic dynamics converge to the deterministic model as the population size grows to infinity, that is, $\lim_{n \to \infty} S_n / N = S$ with constant ϵ fraction of initial infecteds [3].

The algorithm proposed by Molloy and Reed [22] was used to generate the random networks in these experiments. Subsequent research has shown that imperfections can arise in the networks generated by this algorithm, but such biases should be tolerably small for these purposes [21].

The simulation dynamics are as follows:

- A random network of size N is generated.
- A node is chosen uniformly at random from the network as an initial infected.
- An infected node v will recover after an exponentially distributed random time interval $\Delta t_\mu \sim Exp(\mu)$.
- When a node v is infected, each arc (v, x) has a time of infection Δt_x drawn from an exponential distribution $Exp(r)$. If $\Delta t_x < \Delta t_\mu$, node x is infected after time Δt_x. Otherwise x is not infected by v.

This process continues until there are no more infectious nodes.

Figure 4 shows the results of 450 simulations for the Poisson random network considered in the last section ($z = 3$) with 10^4 nodes. The black dotted line represents
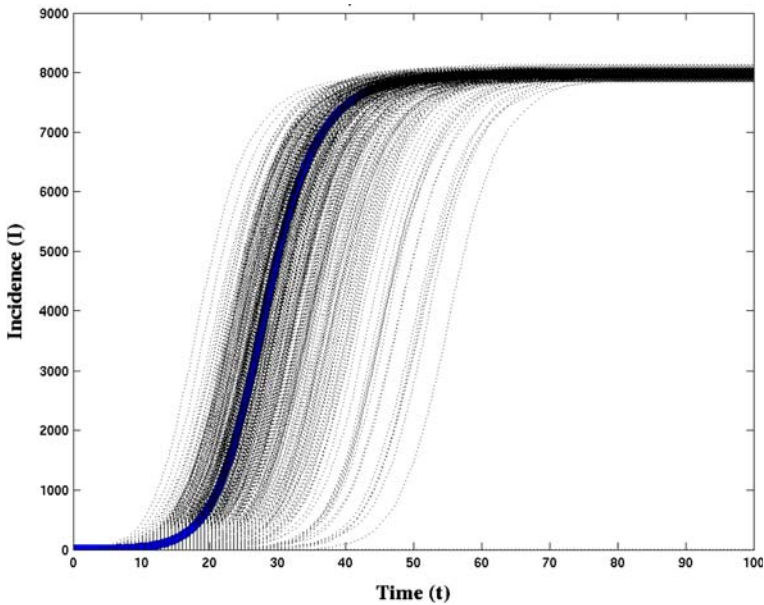
Fig. 4 450 simulation trajectories of the cumulative epidemic incidence J (dotted lines) for a Poisson ($z = 3$) random network with $N = 10^4$ nodes. The *solid line* shows the analytical solution ($\epsilon = 10^{-4}$)

an independent simulation trajectory. The thick, solid line that cuts through the dense mass of simulation trajectories is the analytical trajectory based on the equations in Table 3. The initial conditions were chosen as in the previous section using $\epsilon = 10^{-4}$.

Figure 5 shows a similar series of simulations for the power law degree distribution considered in the last section. In both cases, the analytical trajectory traverses the region with the highest density of simulation trajectories. The simulation trajectories also exhibit significant variability in the time required to reach the expansion phase and final size. This is largely due to the significant impact of random events early on in the epidemic. For example, an initial infected with a low average degree, or one which takes an inordinate amount of time to infect the next infected can markedly delay the onset of the expansion phase.

Figure 6 shows the median-time incidence for the exponential and Poisson networks discussed in the last section. The data points show the median time required to reach a given cumulative incidence among 450 simulation trajectories. The solid line shows the analytical trajectory based on the system of equations given in Table 3. Intuitively, the data points are showing the path of the most central trajectory from the swarm of simulation trajectories such as in Fig. 4.

4 Discussion

The statistical properties of SIR epidemics in random networks have been understood for some time, but the explicit dynamics have been understood mainly through simulation. This paper has addressed this shortcoming by proposing a system of nonlinear ordinary differential equations to model SIR dynamics in random networks.
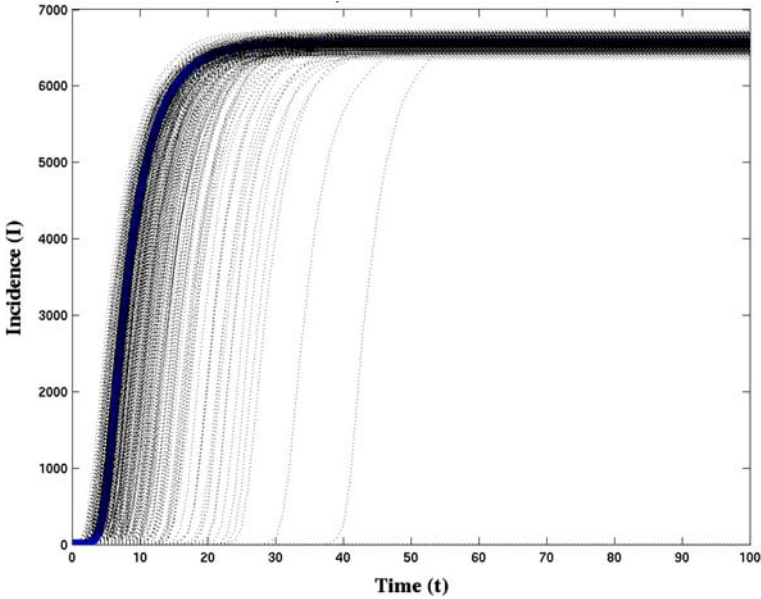
Fig. 5 Four hundred and fifty simulation trajectories of the cumulative epidemic incidence J for a power law ($\gamma = 1.615, \kappa = 20$) random network. $N = 10^4, \epsilon = 10^{-4}$. The *solid line* shows the analytical solution based on the system of equations in Table 3
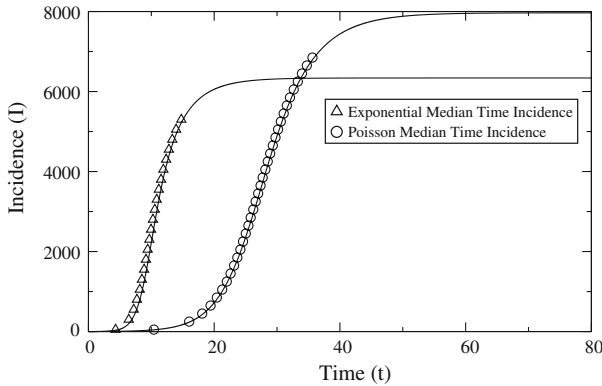
Fig. 6 The median simulation time required to reach a given cumulative incidence J is shown for a Poisson network ($z = 3$, *circles*) and an exponential network ($\lambda = 3.475$, *triangles*). $N = 10^4$, $\epsilon = 10^{-4}$. The *solid line* shows the analytical solution based on the system of equations in Table 3

It should be noted that the SI dynamics are a special case of this model ($\mu = 0$), in which case the ultimate extent of the epidemic is simply the giant component[5] of the network.

---

[5] The *giant component* of a network, if it exists, is a set of nodes such there exists a path between any two of nodes, and furthermore occupies a non-zero fraction of the network in the limit as network size goes to infinity.

The distribution of contacts, even holding the density of contacts constant, has enormous impact on epidemic behavior. This goes beyond merely the extent of the epidemic, but as shown here, the dynamical behavior of the epidemic. In particular, the distribution of contacts plays a key role in determining the onset of the expansion phase.

The distribution dynamics from Eq. 13 and shown in Fig. 3 have important implications for vaccination strategies. Previous work [14, 16] has focused on determining the critical levels of vaccination required to halt or prevent an epidemic. It is usually taken for granted that contact patterns among susceptibles are constant. Furthermore, most widespread vaccinations occur only once an epidemic is underway. Future research could be enhanced by considering optimal vaccination levels when the epidemic proceeds unhindered for variable amounts of time.

It is hoped that the distribution dynamics described in this paper will find applications beyond modeling heterogeneous connectivity. The dynamic PGF approach might be used to capture other forms of heterogeneity, such as of susceptibility, mortality, and infectiousness.

References

1. Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford (1991)
2. Andersson, H.: Epidemic models and social networks. Math. Sci. **24**, 128–147 (1999)
3. Andersson, H., Britton, T.: Stochastic Epidemic Models and their Statistical Analysis. Springer, Heidelberg (2000)
4. Athreya, K.B., Ney, P.: Branching Processes. Springer, New York (1972)
5. Barthelemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A.: Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. J. Theor. Biol. **235**, 275–288 (2005)
6. Bauch, C.T.: A versatile ODE approximation to a network model for the spread of sexually transmitted diseases. J. Math. Biol. **45**(5), 375–395 (2002)
7. Boguna, M., Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in complex networks with degree correlations. In: Rubi, J.M. et al. (ed.), Statistical Mechanics of Complex Networks, Berlin. Springer, Heidelberg (2003)
8. Dezso, Z., Barabasi, A.L.: Halting viruses in scale-free networks. Phys. Rev. E **65**, 055103(R) (2002)
9. Diekmann, O., Heesterbeek, J.A.P.: Mathematical epidemiology of infectious diseases. Model building, analysis and interpretation. Wiley, Chichester (2000)
10. Durrett, R.: Random Graph Dynamics. Cambridge University Press, New York (2007)
11. Eames, T.D., Keeling, M.J.: Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. PNAS **99**, 13330–13335 (2002)
12. Eubank, S., Guclu, H., Anil-Kunar, V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic social networks. Nature **429**, 180–184 (2005)
13. Gupta, S., Anderson, R.M., May, R.M.: Networks of sexual contacts: implications for the pattern of spread of hiv. AIDS **3**, 807–817 (1989)
14. Halloran, M.E., Longini, I., Nizam, A., Yang, Y.: Containing bioterrorist smallpox. Science **298**, 1428 (2005)
15. Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)
16. Kaplan, E.H., Craft, D.L., Wein, L.M.: Emergency response to a smallpox attach: the case for mass vaccination. PNAS USA **99**, 10935 (2002)

17. Keeling, M.J.: The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. B Biol. Sci. **266**(1421), 859–859 (1999)
18. Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y.: The web of human sexual contacts. Nature **411**, 907–908 (2001)
19. Meyers, L.A., Pourbohloul, B., Newman, M.E.J., Skowronski, D.M., Brun-ham, R.C.: Network theory and sars: predicting outbreak diversity. J. Theor. Biol. **232**, 71–81 (2005)
20. Meyers, L.A., Pourbohloul, B., Newman, M.E.J., Skowronski, D.M., Brunham, R.C.: Network theory and SARS: predicting outbreak diversity. J. Theor. Biol. **232**(1), 71–81 (2005)
21. Milo, R., Kashtan, N., Itzkovitz, S., Newman, M.E.J., Alon, U.: Uniform generation of random graphs with arbitrary degree sequences. Preprint cond-mat/0312028 (2003)
22. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. Random Struct. Algor. **6**, 161 (1995)
23. Molloy, M., Reed, B.: The size of the giant component of a random graph with a given degree sequence. Comb. Probab. Comput. **7**, 295–305 (1998)
24. Newman, M.E.J.: The spread of epidemic disease on networks. Phys. Rev. E **66**, 016128 (2002)
25. Newman, M.E.J., Barabási, A.L., Watts, D.J.: The Structure and Dynamics of Networks. Princeton University Press, (2006)
26. Newman, M.E.J., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. PNAS USA **99**, 2566–2572 (2002)
27. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett. **86**, 3200–3203 (2001)
28. Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics and endemic states in complex networks. Phys. Rev. E **63**, 066117 (2001)
29. Pastor-Satorras, R., Vespignani, A.: Handbook of Graphs and Networks: From the Genome to the Internet. chapter Epidemics and immunization in scale-free networks. Wiley-VCH, Berlin (2002)
30. Saramki, J., Kaski, K.: Modelling development of epidemics with dynamic small-world networks. J. Theor. Biol. **234**, 413–421 (2005)
31. Strogatz, S.H.: Exploring complex networks. Nature **410**, 268–276 (2001)
32. Veliov, V.M.: On the effect of population heterogeneity on dynamics of epidemic diseases. J. Math. Biol. **51**, 123–143 (2005)
33. Warren, C.P., Sander, L.M., Sokolov, I., Simon, C., Koopman, J.: Percolation on disordered networks as a model for epidemics. Math. Biosci. **180**, 293–305 (2002)
34. Wilf, H.S.: Generatingfunctionology, 2nd edn. Academic, Boston (1994)