ANZIAM J. 54(2012), 3–22
doi:10.1017/S1446181112000302

EPIDEMIC DYNAMICS ON RANDOM AND SCALE-FREE NETWORKS

J. BARTLETT[1] and M. J. PLANK[✉1]

(Received 28 June, 2012; revised 18 September, 2012; first published online 30 January, 2013)

Abstract

Random networks were first used to model epidemic dynamics in the 1950s, but in the last decade it has been realized that scale-free networks more accurately represent the network structure of many real-world situations. Here we give an analytical and a Monte Carlo method for approximating the basic reproduction number R_0 of an infectious agent on a network. We investigate how final epidemic size depends on R_0 and on network density in random networks and in scale-free networks with a Pareto exponent of 3. Our results show that: (i) an epidemic on a random network has the same average final size as an epidemic in a well-mixed population with the same value of R_0; (ii) an epidemic on a scale-free network has a larger average final size than in an equivalent well-mixed population if $R_0 < 1$, and a smaller average final size than in a well-mixed population if $R_0 > 1$; (iii) an epidemic on a scale-free network spreads more rapidly than an epidemic on a random network or in a well-mixed population.

2010 *Mathematics subject classification*: 92D40.

Keywords and phrases: degree distribution, Pareto distribution, power law, random graph, SIR model, superspreaders.

1. Introduction

The world depends on a wide variety of complex networks, for example social networks [17], the internet [2], networks of ecological interactions [32], biochemical networks [1] and the neurovascular structure of the human brain [10]. It has long been realized that an understanding of the nature and interactions of these networks can advance our understanding of the systems that operate on them. Until about a decade ago, study in this field was largely based on the assumption built into the work of Erdős and Rényi [14] that these networks are randomly connected, with Poisson degree distribution [21]. More recently, it has come to light that more complex structures are common in numerous contexts, such as those mentioned above. Furthermore, it has been realized that many of these share two properties.

[1]Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand; e-mail: michael.plank@canterbury.ac.nz.

The first is the small-world property, which means that although the probability of two randomly selected nodes being connected is extremely low, the number of steps required to get from one node to any other node is relatively small [39]. This property was popularized by the idea that you are six handshakes away from anyone in the world [38].

The second property is a scale-free distribution of degrees. The degree of a node is the number of other nodes to which it is connected. A scale-free distribution means that the probability that a randomly chosen node has degree k is given by a power law or Pareto distribution:

$$P(k) = ck^{-\lambda}, \quad k = 1, 2, 3, \ldots,$$

where $1 < \lambda \leq 3$ and c is a normalization constant [30]. This distribution is heavy-tailed, meaning that it does not have finite variance. While most nodes will have relatively low degree, nodes of arbitrarily large degree will occur in the large-population limit. Even in finite populations, nodes with very large degree occur; these nodes are referred to as hubs. Hubs are critical to the structure of the network, since removing even a few of them can cause the network to become disconnected. Barabási and Albert [5] developed an algorithm for constructing scale-free networks with $\lambda = 3$. Their algorithm incorporates two key requirements for a scale-free network: growth of the network and preferential attachment (meaning that new nodes are more likely to connect to existing nodes that have many connections than to nodes with few connections).

Since the introduction of the classical Kermack–McKendrick model [22], much research has been directed towards producing effective models of disease spread. The usual approach is to divide the population into several classes. At the simplest level, these classes are the members of the population who are susceptible to the disease (S), those who are infectious (I), and those who are removed (R), meaning that they play no further role in the epidemic. It is possible to use additional classes, for example members of the population who have been infected with the disease (exposed) but are not yet infectious to others, or members of the population who have been placed in quarantine [23]. Some models do not use the removed class, assuming instead that infectious individuals remain infectious until the end of the epidemic [6]. Another variant is the susceptible–infectious–susceptible model, in which recovered individuals do not gain immunity, but return to the susceptible class. This allows a stable endemic equilibrium to be reached, in which the proportion of the population in the infectious class reaches a statistically stationary state [18, 40]. Other extensions to the basic susceptible–infectious–removed (SIR) model include dividing the population into age classes [11, 36], incorporating partial immunity [4], natural births, deaths and migration of individuals [8], and investigating the effects of various control strategies [35, 40].

Classical epidemic models assume that the population is homogeneous, meaning that all individuals in a given class are essentially identical [8, 22, 33]. The state of the system is fully described by the size of each class, and the dynamics are usually

described by a system of ordinary [22] or integro-differential [33] equations for the class sizes. More recently, research has been directed towards understanding the effects of network structure on the dynamics of these models. Instead of assuming that the population is well mixed (that is, that an infectious individual can infect any member of the susceptible class), network models assume that the disease can only be transmitted between connected nodes. This means that the state of the system cannot be described by the class sizes alone and more complex approaches are required, such as tracking the numbers of connected pairs of nodes in all possible combinations of classes (S–S, S–I, and so on) [23] or using individual-based simulations.

In a large, well-mixed population, it is frequently possible to obtain analytical expressions for the basic reproduction number R_0 [3, 11, 34]. In a network model, this can be more difficult, and we show that a naïve expression for R_0 can be inaccurate when the mean number of contacts (the mean degree) is relatively low. We instead give a Monte Carlo method for approximating R_0, which can be used for any network with a known degree distribution.

Over the last decade or so, researchers have begun to investigate how epidemic dynamics are affected by a scale-free network structure. It is known that scale-free networks reduce the threshold population size needed for an epidemic to spread [24, 29] and that the spread rate becomes instantaneous in the large-population limit [6]. The presence of hubs makes scale-free networks robust to the removal of large numbers of randomly selected nodes, but vulnerable to the targeted removal of a few highly connected nodes [9, 13, 25]. This means that immunization of randomly selected individuals is far less effective in reducing epidemic size than targeted immunization of highly connected individuals, sometimes referred to as superspreaders [19, 24, 27, 28]. Kiss et al. [24] investigated the effects of contact tracing in a scale-free network. They used a network generation algorithm that assigns each node a degree from an exponentially truncated Pareto distribution $P(k) = ck^{-\lambda}e^{-k/L}$ with $\lambda = 2.5$, $L = 100$ and a fixed mean degree of 6. However, the exponent λ is not typically known for real networks and the effect of choosing different values of λ has not been investigated. Here, we use the Barabási–Albert algorithm [5], which gives a Pareto degree distribution with $\lambda = 3$. (The distribution is naturally truncated as no node can have degree greater than the size of the network.) We simulate a simple SIR model on these networks and compare the epidemic dynamics to random networks and to a well-mixed population. We investigate the effect of varying the basic reproduction number R_0 and the mean number of connections per node (the mean degree).

2. Models

Consider a network consisting of a fixed number of nodes N described by a binary $N \times N$ matrix A: $A_{ij} = 1$ if node i is connected to node j and $A_{ij} = 0$ otherwise. (Clearly A must be symmetric, namely $A_{ij} = A_{ji}$ for all $i, j = 1, \ldots, N$.) If $A_{ij} = 1$, node i is said to be a neighbour, or contact, of node j. A node's degree is its number of contacts. We model the spread of an infectious disease across this network using an individual-based SIR model [11].

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

2.1. Epidemic model At a given point in time, each node is in one of three classes: susceptible, infectious or removed. The numbers of nodes in each of these classes at time t are denoted by $S(t)$, $I(t)$ and $R(t)$, respectively. We use a simple SIR model based on Poisson processes for infection and removal events, as follows:

(i) A susceptible node has a constant probability per unit time β of becoming infected by a neighbouring infectious node. The parameter β is referred to as the transmission coefficient.

(ii) An infectious node has a constant probability per unit time γ of becoming removed. The parameter γ is referred to as the removal rate.

(iii) A removed node plays no further part in the epidemic and remains in the removed class for all time.

These assumptions contain some obvious simplifications. For instance, once infected, an individual is considered to be immediately infectious; no latency period is included in the model. The total population size (including removed nodes) $S(t) + I(t) + R(t)$ is assumed to be constant, so births, deaths, immigration and emigration are ignored over the time scale of the epidemic. For real infectious agents, the probabilities per unit time of infection and removal are not constant, but depend on the time elapsed since infection [26, 33]. A susceptible node's probability of becoming infected per unit time is proportional to its number of infectious contacts i. Some previous models ignore the effect of multiple infectious contacts, instead setting a fixed probability per unit time of infection that is independent of i (provided that $i \geq 1$) [7, 13].

A crucial idea in models of this type is that of the basic reproduction number R_0, defined as the expected number of secondary infections arising from a single infected individual in an otherwise fully susceptible population [11]. That is to say, R_0 is the expected total number of transmissions by a particular infective if all its contacts are susceptible. Naturally, $R_0 = 1$ is a threshold value: if R_0 is less than 1 then the outbreak will rapidly die out, whereas larger values of R_0 have the potential to lead to a major outbreak.

In a finite population, the epidemic will always die out eventually since the number of susceptibles will eventually become small enough that new transmissions are extremely rare. The epidemic is said to be over when $I(t) = 0$. A useful measure of the significance of an outbreak is its final size, which is defined as the total proportion of the population that contracts the disease. This is equivalent to the proportion of the population in the removed compartment once the epidemic is over: $1 - S_\infty/N$, where $S_\infty = \lim_{t\to\infty} S(t)$.

2.2. Network models We consider three types of network model: a well-mixed population, a random network, and a scale-free network. In a well-mixed population, every node is connected to every other node: $A_{ij} = 1$ for all $i, j = 1, \ldots, N$. For a large population size N, the dynamics of the epidemic can be described by the well-known Kermack–McKendrick [22] SIR equations:

$$\frac{dS}{dt} = -\beta IS, \quad \frac{dI}{dt} = \beta IS - \gamma I, \quad \frac{dR}{dt} = \gamma I. \tag{2.1}$$

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

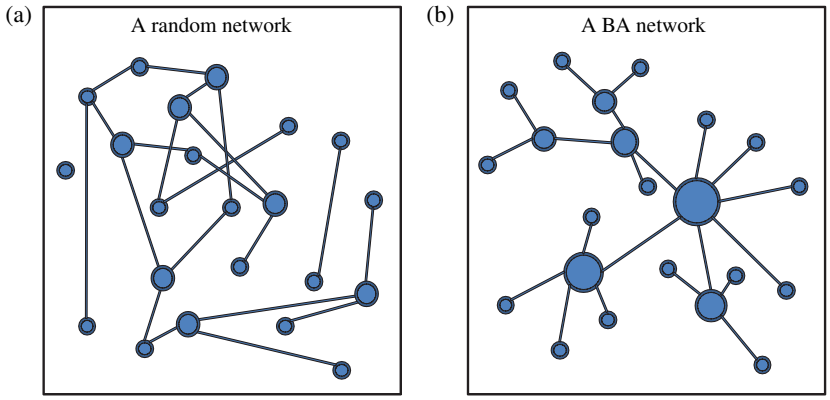


FIGURE 1. Example networks of size $N = 22$: (a) a random network with density $K = 0.1$, (b) a network generated using the BA algorithm with $m = 1$. To help illustrate the degree distribution, the size of each node is proportional to its degree. These example networks are for illustrative purposes; the networks used in epidemic simulations are much larger than this.

The basic reproduction number in a large, well-mixed population is given by

$$R_0 = \frac{\beta N}{\gamma} \tag{2.2}$$

and the final size satisfies the following equation [11]:

$$\ln\left(\frac{S_\infty}{N}\right) = \ln\left(\frac{S(0)}{N}\right) + R_0\left(\frac{S_\infty}{N} - 1\right). \tag{2.3}$$

In a random network, each pair of nodes is connected with fixed probability K, independently of all other pairs of nodes. The parameter K is referred to as the network density. This construction, often termed a random graph, is simple to implement, one reason why it was favoured in epidemic modelling for the four decades following the work of Erdős and Rényi [14]. While the construction is straightforward, it has several surprising consequences in the large-population limit. The most immediate is that all countably infinite random graphs are equivalent, regardless of the density parameter K used in their construction [31]. Hence we can think of all random networks constructed as described above as being approximations of the random graph, called the Rado graph or Erdős–Rényi graph [14, 31]. The degree distribution of a random network is binomial, which, for large populations, is well approximated by a Poisson distribution. This implies that the vast majority of nodes have degree lying in a relatively narrow range about the mean degree. Figure 1(a) shows an example of a random network.

Barabási and Albert [5] showed that the combination of network growth and preferential attachment can be used to generate networks that are scale free. Growth means that the network is not static and new nodes are added with time. Preferential attachment means that when a new node is added, it is connected to an existing node

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

with a probability that is proportional to the degree of that node. There is therefore a self-reinforcing tendency for a few nodes to become highly connected, while the majority of nodes have relatively few connections. Although there are other methods of construction [12], we use the Barabási–Albert (BA) algorithm [5] to construct scale-free networks. We begin with a small, fully connected network of size m and then add new nodes in an iterative process. At each iteration, a new node is added and is connected to m randomly selected existing nodes. The probability that the new connection is with node i is

$$p_i = \frac{k_i}{\sum_{j \in S} k_j},$$

where k_i is the current degree of node i and S is the set of all nodes to which the new node is not yet connected. This process is repeated until the network has the desired number of nodes N. This results in a network whose degree distribution approaches an inverse-cubic Pareto distribution

$$P(k) = 2m^2/k^3, \quad m \leq k < \infty, \tag{2.4}$$

as N approaches infinity [5]. The mean of this degree distribution is $2m$ and the variance is infinite. Because of this, there is no "typical" degree and, although most nodes will have relatively small degree, there will be a small number of nodes with very large degree. By construction, the density of the network is

$$K = \frac{m(2N-1) - m^2}{N(N-1)},$$

which is approximately $2m/N$ for large N. The size of the initial fully connected seed group m can thus be chosen to generate a network of any required density $0 \leq K \leq 1$. Figure 1(b) shows an example of a network constructed using the BA algorithm.

Note that the iterative process described above was first used to generate a network and the epidemic model was then run on that static network. This is effectively an assumption that the time scale of the epidemic outbreak is much faster than the birth–death process in the population.

2.3. Numerical methods Infections and removals both occur as Poisson processes with rates $\beta[SI]$ and γI, respectively, where $[SI]$ denotes the number of connected SI pairs (in the well-mixed case, $[SI] = SI$). These processes are simulated using the Gillespie [16] algorithm, as follows. The sum of the infection and removal processes is itself a Poisson process, with rate $\beta[SI] + \gamma I$. Therefore, the times between events are drawn from an exponential distribution: $T \sim \mathrm{Exp}(\beta[SI] + \gamma I)$. Once the event time has been drawn, the type of event is decided. The event is an infection with probability $P_I = \beta[SI]/(\beta[SI] + \gamma)$ and a removal with probability $P_R = 1 - P_I$. In the case of an infection event, the SI pair that the event applies to is drawn at random from the set of all SI pairs. In the case of a removal event, the node to which the event applies is drawn at random from all nodes in the infectious class. The status of the affected

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

nodes is then updated and the process is repeated until there are no infectious nodes remaining.

Equation (2.3) was solved using Newton's method. All computations were performed in MATLAB and the system (2.1) was solved using the built-in ordinary differential equation solver *ode*45. A population size of $N = 500$ or $N = 1000$ was used in all simulations. This allowed us to investigate some of the finite-size effects that are important in the network models studied. Using a larger population, for the same network density K, shifts the results towards the infinite-population limit. The simulation method was verified by comparing simulations of the well-mixed model to solutions of the Kermack–McKendrick SIR equations (2.1) and to simulations of a fully connected network. Averaging over 50 realizations, these results matched up almost exactly (for example, see Figure 4). Including more realizations slightly reduced the amount of noise in Figures 2, 4 and 7, but did not change any of the overall patterns.

3. Results

3.1. Random networks We first investigate how network density affects the final size of an epidemic on a random network. Final size was calculated by averaging $1 - S_\infty$ over n realizations of the individual-based epidemic model. In a large population, R_0 is equal to β/γ multiplied by the number of potential secondary cases, which, in the well-mixed case, is simply the population size N [11]. For a random network, the number of potential secondary cases is the degree of the index node (the primary infection). The mean degree is approximately NK, so

$$R^* = \frac{\beta NK}{\gamma} \tag{3.1}$$

might be expected to provide a good predictor final size. Figure 2 shows final size against R^* for a range of network densities K and transmission coefficients β on a random network of size $N = 500$. Network density K does not have a major impact on final size for K greater than about 0.1, and final size is accurately predicted by equation (2.3) (Figure 2, bottom panels). As K drops below 0.1, however, final size decreases substantially for a fixed transmission coefficient β (Figure 2, top panels).

Equation (2.2) for R_0 is only valid when the number of potential secondary cases is large, so that the rate of new infections by the primary case does not drop substantially as the primary case's contacts move from the susceptible to the infectious class [11]. For the small values of network density in Figure 2, the primary case has a relatively small number of connections. As a consequence, the potential for new infections drops substantially each time the primary case infects one of its contacts.

Hence, R^* is not a good measure of the epidemic's potential when the mean degree is small. A more accurate approximation for R_0 may be obtained by noting that a node exposed to the primary case for a period of time τ remains uninfected with probability $e^{-\beta\tau}$. Suppose that the primary case has k_0 contacts and remains infectious for a period

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

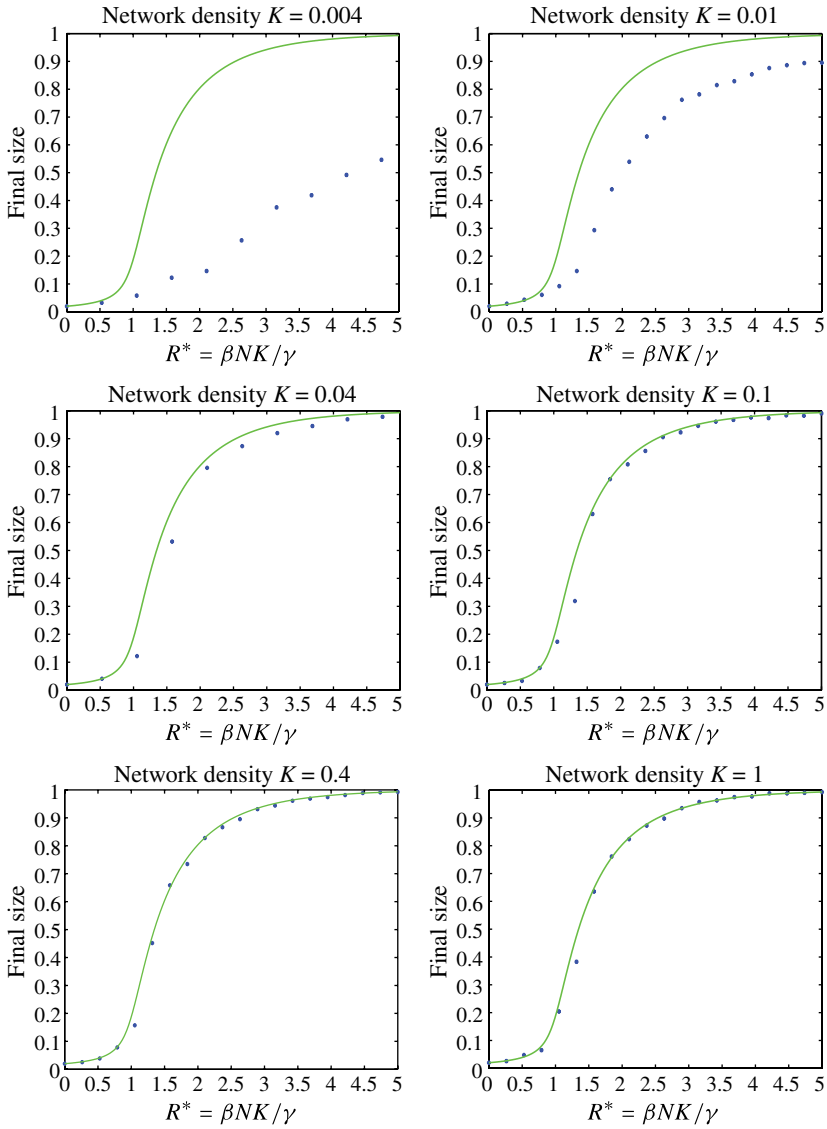


FIGURE 2. Final size as a function of $R^* = \beta NK/\gamma$ in random networks. In each panel, R^* was varied by varying β and keeping $N = 500$, $\gamma = 1$ and K fixed. Points represent the average final size from 10 realizations of the individual-based model. Simulations were initialized with a randomly chosen 2% of nodes being infectious and all remaining nodes being susceptible. The curve is the numerical solution to equation (2.3) with $S(0) = 0.98N$. This gives a good prediction of final size only when network density K is greater than about 0.1, but overestimates final size when K is less than 0.1.

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

of time τ. We approximate the number of secondary infections R from the primary case by a binomial random variable $R \sim \mathrm{Bin}(k_0, 1 - e^{-\beta\tau})$, which has expected value $k_0(1 - e^{-\beta\tau})$. Conditioning on the values of k_0 and τ gives

$$R_0 = \int_0^\infty f_I(\tau)\Big(\sum_{k_0=0}^{N-1} P(k_0)k_0(1 - e^{-\beta\tau})\Big)\, d\tau = \bar{k}\Big(1 - \int_0^\infty f_I(\tau)e^{-\beta\tau}\, d\tau\Big), \tag{3.2}$$

where f_I is the probability density function for the infectious period, $P(k)$ is the probability that a randomly chosen node has k contacts, and $\bar{k}$ is the mean degree.

In the case of a random network, the mean degree is $\bar{k} = (N - 1)K$. The length of time τ for which the primary case is infectious has an exponential distribution, $f_I(\tau) = \gamma e^{-\gamma\tau}$. Hence

$$R_0 = \frac{\beta(N - 1)K}{\gamma + \beta}.$$

In general, equation (3.2) gives an analytical approximation for R_0 provided that the integral over τ can be evaluated and the infectious period τ, the transmission coefficient β and the recovery rate γ are all independent of the number of contacts k_0. The integral over τ can be evaluated for a reasonably general range of distributions $f_I(\tau)$, for example a fixed infection period, a gamma distribution as considered by Ross [37], or a normal distribution. However, there may be contexts in which the independence conditions do not hold. In such cases, R_0 can instead be estimated using a Monte Carlo method. Generating random values for the number of contacts k_0, infectious period τ and number of secondary infections R according to

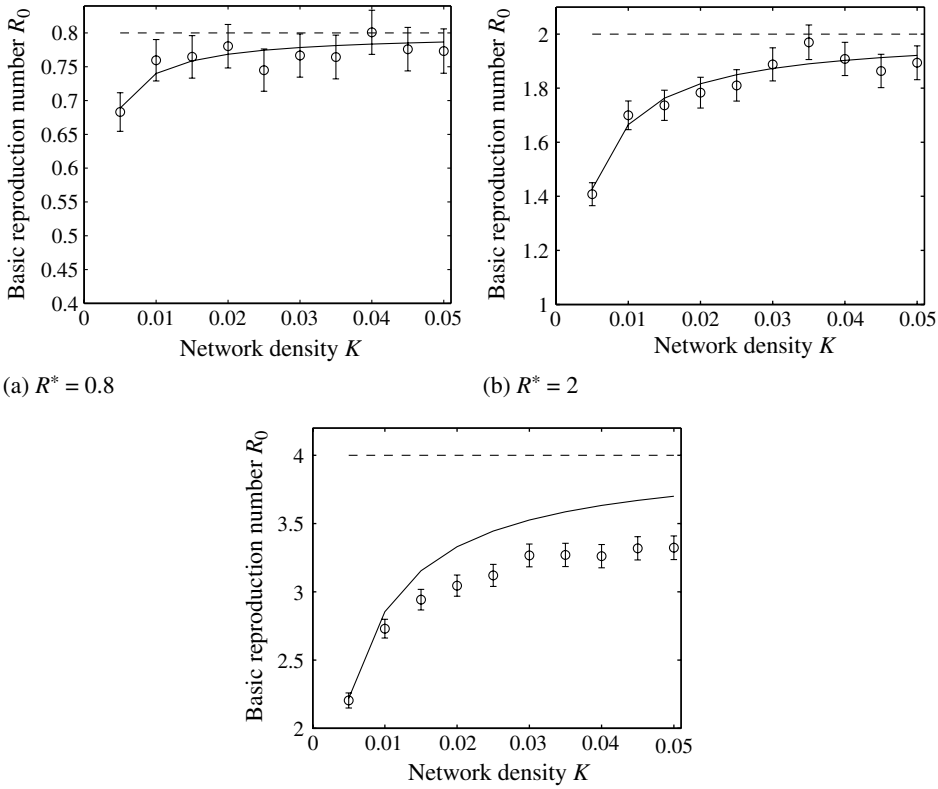
$$\tau \sim \mathrm{Exp}(\gamma), \quad k_0 \sim \mathrm{Bin}(N - 1, K), \quad R \sim \mathrm{Bin}(k_0, 1 - e^{-\beta\tau}) \tag{3.3}$$

gives a random sample of R. Repeating this process a sufficiently large number of times and calculating the mean of R gives an estimate for R_0.

Both equation (3.2) and the Monte Carlo method (3.3) provide approximations to the true value of R_0. They both neglect the possibility that one of the primary case's contacts could be infected by another node while the primary case is still infectious, rather than by the primary case directly. This assumption is likely to be more serious in clustered networks, where the primary case's contacts are more likely to be connected to one another. A method for well-mixed populations that avoids this assumption is given by Ross [37].

Figure 3 compares the naïve approximation (3.1) and the improved approximation (3.2) to the value of R_0 obtained from simulations of the full network model. (Values of R_0 generated using the Monte Carlo method (3.3) are almost identical to those calculated using equation (3.2).) These graphs show that equation (3.2) provides a much more accurate approximation to R_0 than the naïve approximation. When R_0 is subcritical (Figure 3(a)) or moderately supercritical (Figure 3(b)), the improved approximation lies within the 95% confidence interval for R_0 in almost all cases. Equation (3.2) loses accuracy when R_0 is high (Figure 3(c)). This is because of the

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press



(a) $R^* = 0.8$ (b) $R^* = 2$

(c) $R^* = 4$

FIGURE 3. Comparison of the naïve (dashed line) and improved (solid curve) approximations to R_0 with values of R_0 calculated from $m = 1000$ repeat simulations of the full network model (points) for varying network density K and three different values of R^*. The vertical axis ranges from $0.5R^*$ to $1.05R^*$ in each case. Points show the mean number of nodes $\bar{R}$ infected by the primary case during the simulation. Error bars show 95% confidence intervals for R_0, calculated as $\bar{R} \pm 1.96\sigma/\sqrt{m}$. Parameter values: $N = 1000$, $\gamma = 1$, with β varied according to $\beta = \gamma R^*/(NK)$ to give the required value of R^*.

increased likelihood of contacts of the primary case being infected by other nodes while the primary case is still infectious. However, equation (3.2) is still much more accurate than equation (3.1).

An approximate value for R_0 can be obtained very quickly using either equation (3.2), when available, or the Monte Carlo method (3.3). Calculating R_0 from repeat simulations of the full network model is much more computationally intensive.

Figure 4 shows final size against the value of R_0 calculated from (3.3). The results shows that, when R_0 is calculated using this method, it provides a good predictor of final size regardless of the network density.

3.2. Scale-free networks In order to check that networks generated using the BA algorithm are approximately scale free, we first calculate the degree distribution of a

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

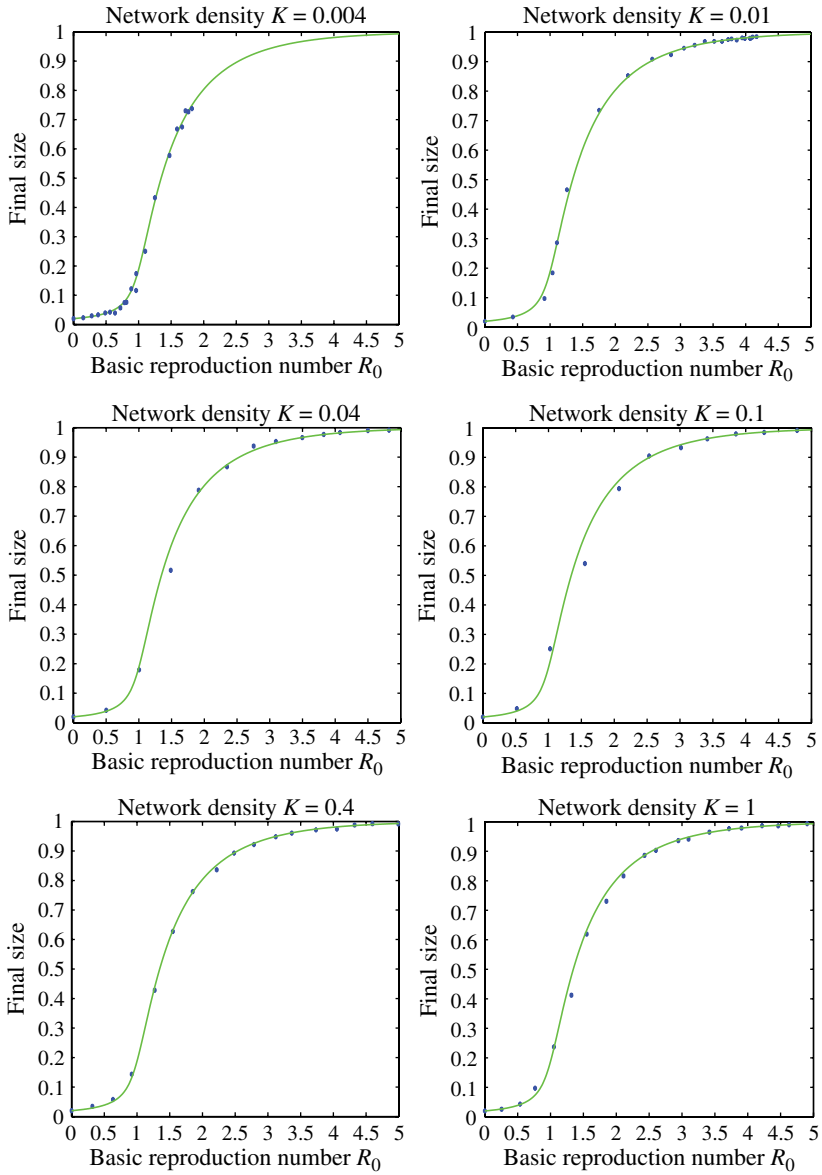


FIGURE 4. Final size as a function of R_0 in random networks. In each panel, R_0 was varied by varying β and keeping $N = 500$, $\gamma = 1$ and K fixed; R_0 was calculated from a Monte Carlo sample of size $m = 10^4$. Blue points represent the average final size from 10 realizations of the individual-based model. Simulations were initialized with a randomly chosen 2% of nodes being infectious and all remaining nodes being susceptible. The green curve is the numerical solution to equation (2.3) with $S(0) = 0.98N$. This gives a good prediction of final size for all network densities investigated. Note that when $K = 0.004$, the mean degree is equal to 2, so it is impossible for R_0 to be greater than 2.

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

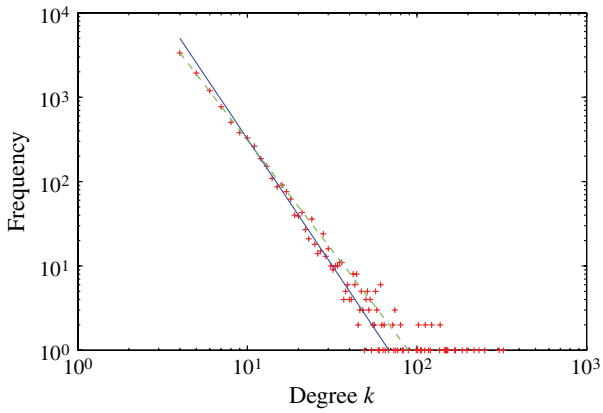


FIGURE 5. Degree distribution (log–log frequency plot) of a network of size $N = 10^4$ generated using the BA algorithm with $m = 4$ (red crosses) and the theoretical Pareto distribution in equation (2.4) (blue line). The green dashed line is the line of best fit for the data.

network of size $N = 10^4$ generated using this method. This compares well with the expected distribution in equation (2.4) (Figure 5).

The value of R_0 for a scale-free network may be calculated using equation (3.2) or with the Monte Carlo method. The number of contacts k_0 of the primary case is distributed approximately according to equation (2.4). However, equation (2.4) applies only in the large-population limit and, while the approximation is reasonably accurate for $N = 10^4$, it may be less accurate for the smaller networks used in the epidemic simulations. To minimize the error in the approximated value for R_0, we therefore sampled values of k_0 by generating a network of the required size using the BA algorithm and counting the number of contacts of a randomly selected node in that network. As for random networks, equation (3.2) and the Monte Carlo method provided near-identical results.

We now compare epidemic dynamics on scale-free networks with those on random networks of the same density. Figure 6 shows time series for epidemics on networks of density $K = 0.01$ and for three different values of R_0. In all cases, the variability among realizations of the stochastic model is greater for scale-free networks than for random networks. This is to be expected because the variance of the degree distribution is much higher, relative to the mean degree, than in a random network. The epidemic also spreads more rapidly and peaks earlier (red $I(t)$ curves in Figure 6) on scale-free networks than on random networks. The final epidemic size (the asymptotic height of the green $R(t)$ curves in Figure 6) is higher for scale-free networks than for random networks when R_0 is close to 1, but lower when R_0 is substantially greater than 1.

Similar comparisons with higher network density reveal that the differences between the random and scale-free networks diminish as network density increases. This is partly because the networks of higher density require a greater value of m in the BA algorithm. As a consequence, a greater proportion of the network is in the initial,

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

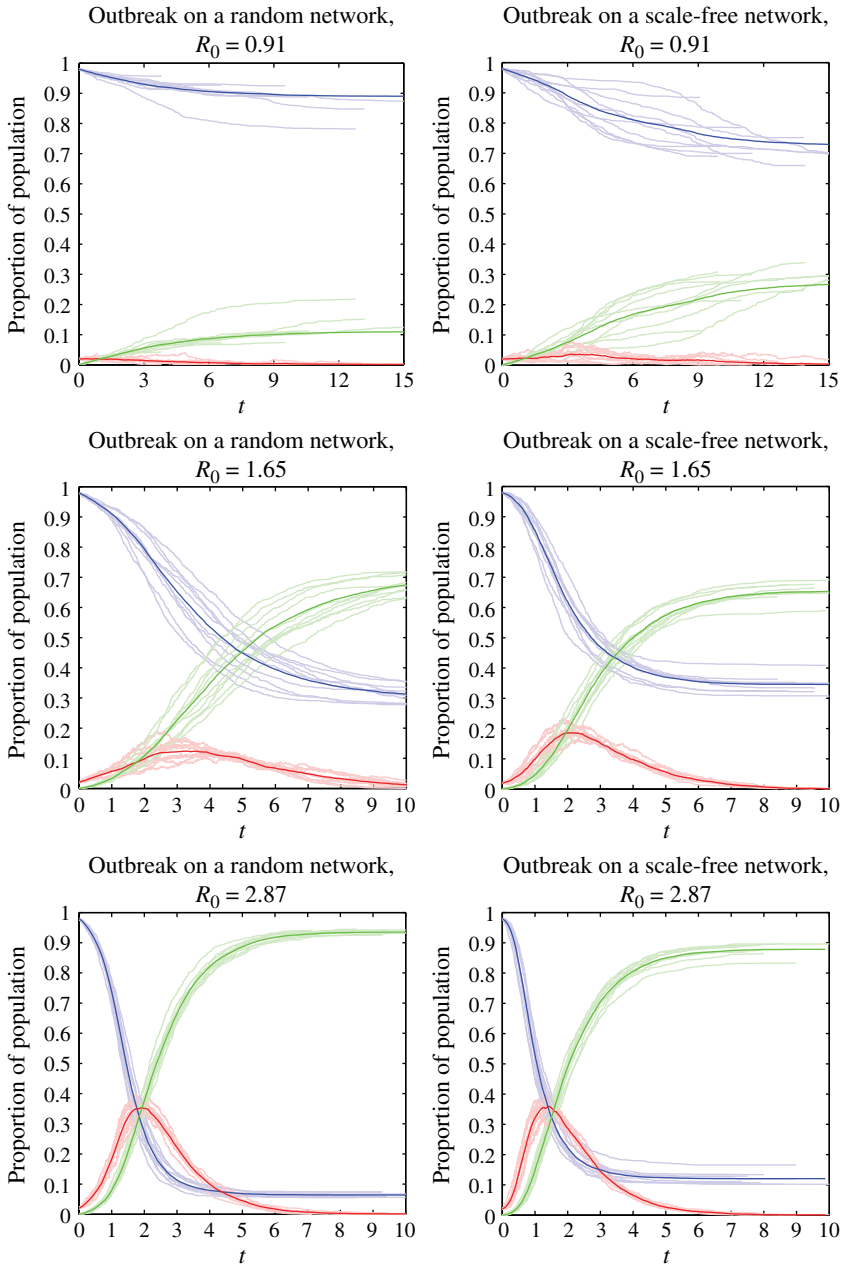


FIGURE 6. Realizations (pale curves) of the stochastic epidemic model on a random network (left) and on a scale-free network (right) of the same size $N = 1000$ and density $K = 0.01$ ($m = 5$ in the BA algorithm): susceptible $S(t)$ (blue), infectious $I(t)$ (red), removed $R(t)$ (green). (Colour available online.) Dark curves show the average of the stochastic realizations. Rates are $\beta = 0.1$ (top row), $\beta = 0.2$ (middle row), $\beta = 0.4$ (bottom row); $\gamma = 1$ in all cases. Time is dimensionless and $t = 1$ corresponds to the mean period of time for which an individual node remains infectious. Note the different time scale in the top graphs to show the full course of the outbreak.

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

fully connected seed group, rather than in the group of nodes whose connections are chosen according to the preferential attachment model. Thus, the degree distribution will be further away from the limiting Pareto distribution, equation (2.4). Furthermore, as network density increases, the limiting distribution itself becomes less distinct from the degree distribution of a random network (the two distributions are trivially identical in the extreme case $K = 1$).

We now investigate the relationship between final size and R_0 in scale-free networks (Figure 7). For large values of R_0, scale-free networks have a lower final size than random networks. However, for smaller values of R_0 close to or less than 1, the opposite effect is seen. This is true for all network densities investigated, though the differences are more pronounced at lower densities for reasons discussed above.

Finally, we investigate how the scale-free network structure affects the time scale of the epidemic. Figure 8 shows the average time between the first infection and the time at which the number of infectious nodes $I(t)$ reaches its peak value. This metric, referred to as time-to-peak, is a useful measure of the spread speed and duration of the epidemic (information that cannot be inferred from the value of R_0 alone). The results show that epidemics progress more rapidly on scale-free networks than on random networks of the same density. This is particularly so for relatively small values of R_0. Note that results for $R_0 < 2$ are not included in Figure 8 because a significant number of realizations of the stochastic model die out very rapidly, which biases the time-to-peak data.

4. Discussion

We have used a simple SIR model to investigate epidemic dynamics in random networks. We have compared three types of network: a well-mixed population, in which an infectious individual can potentially transmit the disease to any other member of the population; a random network, in which all individuals (nodes) have approximately the same number of contacts; and a scale-free network, which is characterized by a heavy-tailed degree distribution, meaning that most nodes have very few contacts, but a small number of nodes (called hubs) have an extremely large number of contacts. We have compared these three models and investigated the effects of varying the basic reproduction number R_0 and the network density (or equivalently the mean degree).

We have shown that care is needed when calculating R_0 in network models where the mean number of contacts per node (the mean degree) is relatively small. Because of the limited pool of potential secondary infections, naïve calculations of R_0 are likely to be inaccurate. Instead, we have provided an analytical approximation for R_0 that is valid in cases where an individual's transmission coefficient and recovery rate are independent of his/her number of contacts. When calculated correctly, R_0 provides an accurate prediction of the proportion of the population that ultimately becomes infected by the disease (the final epidemic size) on a random network. An important implication of this is that an infection in a highly connected population with a low

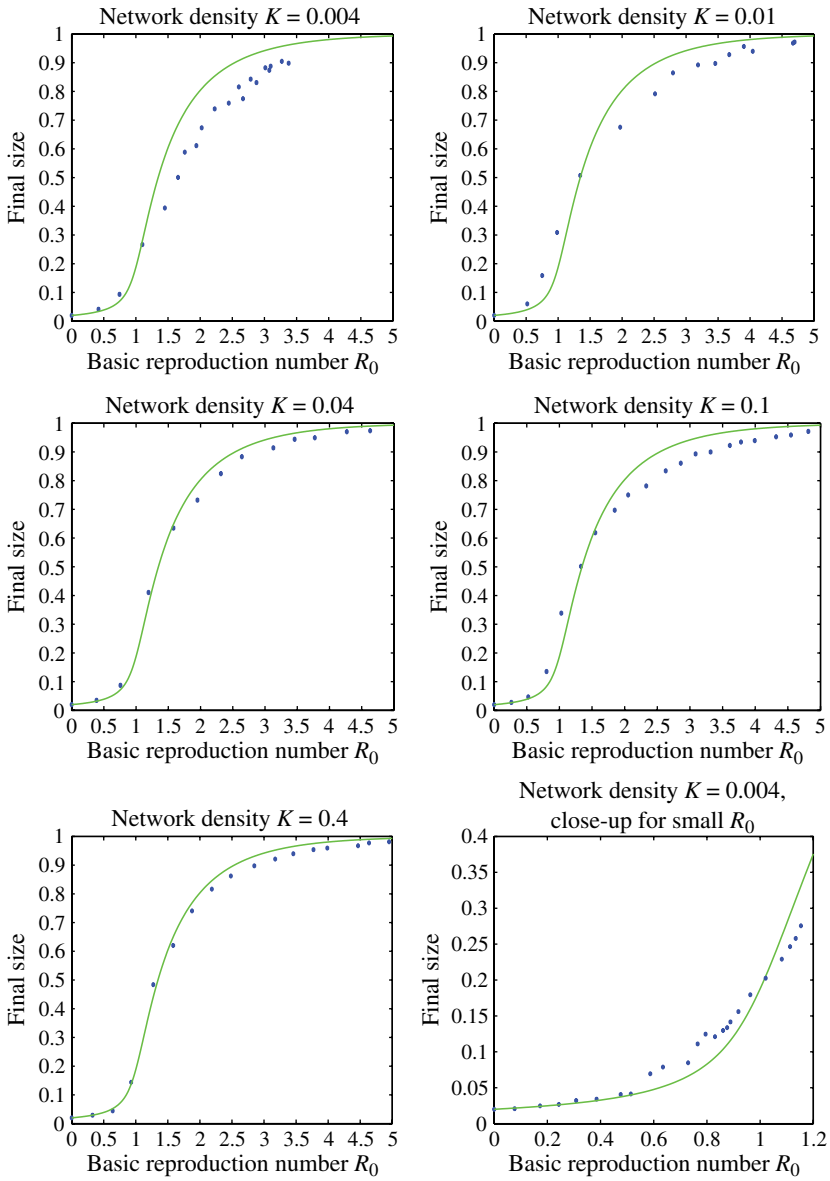


FIGURE 7. Final size as a function of R_0 in scale-free networks. In each panel, R_0 was varied by varying β and keeping $N = 1000$, $\gamma = 1$ and K fixed; R_0 was calculated from a Monte Carlo sample of size $m = 10^4$. Blue points represent the average final size from 10 realizations of the individual-based model. Simulations were initialized with a randomly chosen 2% of nodes being infectious and all remaining nodes being susceptible. The green curve is the numerical solution to equation (2.3) with $S(0) = 0.98N$, which shows the expected final size for a random network. When R_0 is close to or less than 1, the final size is larger on scale-free networks than on random networks of the same density and with the same R_0; when R_0 is much larger than 1, the effect is reversed. The bottom-right panel shows a more detailed close up of the top-left panel for small values of R_0.

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

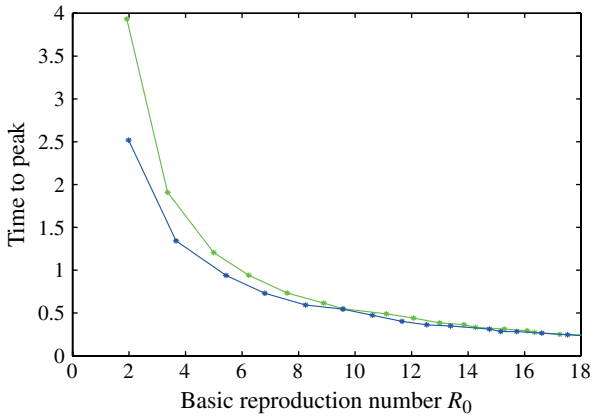


FIGURE 8. Time at which the number of infectious nodes $I(t)$ peaks as a function of R_0 on random (green) and scale-free (blue) networks of size $N = 1000$ and density $K = 0.04$ (colour available online). R_0 is varied by varying β with fixed $\gamma = 1$. Time is dimensionless and $t = 1$ corresponds to the mean period of time for which an individual node remains infectious.

transmission coefficient is expected to have the same final size as an infection in a sparsely connected population with a high transmission coefficient and the same value of R_0.

Our results show that epidemics on scale-free networks initially spread more rapidly than on random networks of the same density. This supports previous results showing that the spread rate of an epidemic on a scale-free network becomes instantaneous in the limit as the network size N tends to infinity [6, 29]. Our results also show that the scale-free network structure can affect the final epidemic size. For values of R_0 close to or less than 1, scale-free networks have a lower final size than random networks; for values of R_0 much greater than 1, the effect is reversed.

These results can be explained by the existence of hubs (or superspreaders), nodes with many more connections than the average. Because of their high connectivity, superspreaders are likely to be infected early in the course of the epidemic, and are then able to pass the disease on to a large number of nodes. As a result, the epidemic can initially spread more rapidly than would be predicted by considering an average node (that is, by R_0). As well as explaining the faster time-to-peak on scale-free networks, this also accounts for the higher final size when R_0 is relatively small. However, when R_0 is larger, the final size is not determined by the initial rapid spread via the superspreaders, but by the ability of the epidemic to infect the poorly connected nodes before dying out. In a random network, almost all nodes have close to the average number of connections. In scale-free networks, there are a large number of poorly connected nodes, which are very isolated and therefore much less likely to become infected. These effects were also observed on scale-free networks (with Pareto exponent $\lambda = 2.5$) by Kiss et al. [24]. However, these authors did not investigate the effect of varying network density, nor calculate the threshold value of R_0 at which the

https://doi.org/10.1017/S1446181112000302 Published online by Cambridge University Press

final size on a scale-free network switches from being higher to lower than that on a random network.

The structure of real networks is rich and varied [30] and we have only a limited understanding of the ways in which this influences the spread and persistence of epidemics. Applications of such understanding are wide-ranging and important, including the potential to protect networks such as the internet, understand complex ecological networks [20] and suggest effective control measures for containing or preventing epidemics of infectious diseases [23, 24]. One potential application is improving vaccination strategies with the aim of preventing, or at least reducing, the final sizes of epidemics. Most nodes in a scale-free network have a relatively low number of contacts. Therefore, randomly selecting members of the population to vaccinate tends to remove these structurally unimportant nodes. Because of this, for an immunization strategy to be effective, a very large proportion of the population needs to be vaccinated. Selectively vaccinating highly connected individuals (superspreaders) is much more effective in preventing an outbreak, but can be impractical or unethical. A suggested compromise is to select a random portion of the population and have each selected individual nominate one of their contacts for vaccination [15, 28]. It is clear that more research is needed to establish effective vaccination strategies for real social networks.

The model presented here could be extended in various ways. We have only considered binary transmission coefficients, that is, one individual can potentially transmit the disease to another individual either at rate β or at rate 0. This could be extended in two ways. Firstly, each individual i might be assigned a shedding rate r_i, representing the probability per unit time that this individual has of transmitting the infection. For example, in a flu epidemic, strategies such as containing coughs and sneezes, taking medication or frequently washing hands may reduce an individual's shedding rate. This would lead to an asymmetric transmission model. Secondly, different pairs of connected individuals could be assigned differing contact rates. The assumption of a constant contact rate means that highly connected superspreaders may infect contacts at unrealistically high rates. Fu et al. [15] investigated the effects of limiting the total infection rate of each individual, but this remains a largely unstudied aspect of network epidemic models. Note that in these more complex cases, an analytical expression for R_0 will not typically be available, but the Monte Carlo method can always be used to estimate R_0.

The fact that outbreaks on scale-free networks were modelled at a fixed endpoint of the construction algorithm of Barabási and Albert [5] suggests a further avenue for analysis. What is the effect of continued growth of the network on the time scale of the outbreak? This would enable the model to incorporate a natural birth–death process while maintaining the scale-free structure of the network.

5. Summary

Our results make several novel contributions to the literature. First, we have provided a method for calculating the basic reproduction number that can be applied

to any network model. Second, we have shown that an epidemic in a random network has the same average final size as an equivalent epidemic (that is, with the same R_0) in a well-mixed population. Third, we have shown that epidemics spread more rapidly on scale-free networks than on random or well-mixed networks. This extends previous results [24] for networks with a Pareto exponent of 2.5 and fixed density. Fourth, we have shown that subcritical epidemics ($R_0 < 1$) have a higher final size on scale-free networks than on random networks, whereas the opposite is true of supercritical epidemics ($R_0 > 1$).

Acknowledgements

James Bartlett thanks James Williams for help with programming the stochastic model and network algorithms. The authors are grateful to Joshua Ross for insightful comments and for pointing out the existence of the analytical approximation for R_0 in a network model.

References

[1] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks", *Rev. Modern Phys.* **74** (2002) 47–97; doi:10.1103/RevModPhys.74.47.

[2] D. Alderson, H. Chang, M. Roughan, S. Uhlig and W. Willinger, "The many facets of internet topology and traffic", *Net. Heterog. Media* **1** (2006) 569–600; doi:10.3934/nhm.2006.1.569.

[3] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control* (Oxford University Press, Oxford, 1991).

[4] V. Andreasen, J. Lin and S. A. Levin, "The dynamics of cocirculating influenza strains conferring partial cross-immunity", *J. Math. Biol.* **35** (1997) 825–842; doi:10.1007/s002850050079.

[5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science* **286** (1999) 509–512; doi:10.1126/science.286.5439.509.

[6] M. Barthélemy, A. Barrat, R. Pastor-Satorras and A. Vespignani, "Velocity and hierarchical spread of epidemic outbreaks in scale-free networks", *Phys. Rev. Lett.* **92** (2004) 178701; doi:10.1103/PhysRevLett.92.178701.

[7] M. Boguñá, R. Pastor-Satorras and A. Vespignani, "Absence of epidemic threshold in scale-free networks with degree correlations", *Phys. Rev. Lett.* **90** (2003) 028701; doi:10.1103/PhysRevLett.90.028701.

[8] F. Brauer and P. van den Driessche, "Models for transmission of disease with immigration of infectives", *Math. Biosci.* **171** (2001) 143–154; doi:10.1016/S0025-5564(01)00057-8.

[9] D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, "Network robustness and fragility: percolation on random graphs", *Phys. Rev. Lett.* **85** (2000) 5468–5471; doi:10.1103/PhysRevLett.85.5468.

[10] T. David, T. van Kempen, H. Huang and P. Wilson, "The geometry and dynamics of binary trees", *Math. Comput. Simulation* **81** (2011) 1464–1481; doi:10.1016/j.matcom.2010.04.020.

[11] O. Diekmann and J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation* (John Wiley & Sons, Chichester, 2000).

[12] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, "Structure of growing networks with preferential linking", *Phys. Rev. Lett.* **85** (2000) 4633–4636; doi:10.1103/PhysRevLett.85.4633.

[13] V. M. Eguíluz and K. Klemm, "Epidemic threshold in structured scale-free networks", *Phys. Rev. Lett.* **89** (2002) 108701; doi:10.1103/PhysRevLett.89.108701.

[14] P. Erdős and A. Rényi, "On random graphs I", *Publ. Math. Debrecen* **6** (1959) 290–297.

[15] X. Fu, M. Small, D. M. Walker and H. Zhang, "Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization", *Phys. Rev. E* **77** (2008) 036113; doi:10.1103/PhysRevE.77.036113.

[16] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", *J. Phys. Chem.* **81** (1977) 2340–2361; doi:10.1021/j100540a008.

[17] L. Hufnagel, D. Brockmann and T. Geisel, "Forecast and control of epidemics in a globalized world", *Proc. Natl. Acad. Sci.* **101** (2004) 15124–15129; doi:10.1073/pnas.0308344101.

[18] J. A. Jacquez and C. P. Simon, "The stochastic SI model with recruitment and deaths I. Comparison with the closed SIS model", *Math. Biosci.* **117** (1993) 77–125; doi:10.1016/0025-5564(93)90018-6.

[19] A. James, J. W. Pitchford and M. J. Plank, "An event-based model of superspreading in epidemics", *Proc. R. Soc. Lond. B* **274** (2007) 741–747; doi:10.1098/rspb.2006.0219.

[20] A. James, J. W. Pitchford and M. J. Plank, "Disentangling nestedness from models of ecological complexity", *Nature* **487** (2012) 227–230; doi:10.1038/nature11214.

[21] M. J. Keeling, "The effects of local spatial structure on epidemiological invasions", *Proc. R. Soc. Lond. B* **266** (1999) 859–867; doi:10.1098/rspb.1999.0716.

[22] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. R. Soc. Lond. A* **115** (1927) 700–721; doi:10.1098/rspa.1927.0118.

[23] I. Z. Kiss, D. M. Green and R. R. Kao, "Disease contact tracing in random and clustered networks", *Proc. R. Soc. Lond. B* **272** (2005) 1407–1414; doi:10.1098/rspb.2005.3092.

[24] I. Z. Kiss, D. M. Green and R. R. Kao, "Infectious disease control using contact tracing in random and scale-free networks", *J. Roy. Soc. Interface* **3** (2006) 55–62; doi:10.1098/rsif.2005.0079.

[25] J.-G. Liu, Z.-T. Wang and Y.-Z. Dang, "Optimization of robustness of scale-free network to random and targeted attacks", *Modern Phys. Lett. B* **19** (2005) 785–792; doi:10.1142/S0217984905008773.

[26] A. L. Lloyd, "Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics", *Theor. Populat. Biol.* **60** (2001) 59–71; doi:10.1006/tpbi.2001.1525.

[27] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp and W. M. Getz, "Superspreading and the effect of individual variation on disease emergence", *Nature* **438** (2005) 355–359; doi:10.1038/nature04153.

[28] N. Madar, T. Kalisky, T. Cohen, D. ben-Avraham and S. Havlin, "Immunization and epidemic dynamics in complex networks", *Eur. Phys. J. B* **38** (2004) 269–276; doi:10.1140/epjb/e2004-00119-8.

[29] R. M. May and A. L. Lloyd, "Infection dynamics on scale-free networks", *Phys. Rev. E* **64** (2001) 066112; doi:10.1103/PhysRevE.64.066112.

[30] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law", *Contemp. Phys.* **46** (2005) 323–351; doi:10.1080/00107510500052444.

[31] R. Rado, "Universal graphs and universal functions", *Acta. Arith.* **9** (1964) 331–340.

[32] E. L. Rezende, J. E. Lavabre, P. R. Guimarães, P. Jordano and J. Bascompte, "Non-random coextinctions in phylogenetically structured mutualistic networks", *Nature* **448** (2007) 925–928; doi:10.1038/nature05956.

[33] M. G. Roberts, "A Kermack–McKendrick model applied to an infectious disease in a natural population", *Math. Med. Biol.* **16** (1999) 319–332; doi:10.1093/imammb16.4.319.

[34] M. G. Roberts, "The pluses and minuses of $\mathcal{R}_0$", *J. Roy. Soc. Interface* **4** (2007) 949–961; doi:10.1098/rsif.2007.1031.

[35] M. G. Roberts and J. A. P. Heesterbeek, "A new method for estimating the effort required to control an infectious disease", *Proc. R. Soc. Lond. B* **270** (2003) 1359–1364; doi:10.1098/rspb.2003.2339.

[36] M. G. Roberts and M. I. Tobias, "Predicting and preventing measles epidemics in New Zealand: application of a mathematical model", *Epidemiol. Infect.* **124** (2000) 279–287; doi:10.1017/S0950268899003556.

[37] J. V. Ross, "Invasion of infectious diseases in finite homogeneous populations", *J. Theoret. Biol.* **289** (2011) 83–89; doi:10.1016/j.jtbi.2011.08.035.

[38] J. Travers and S. Milgram, "An experimental study of the small world problem", *Sociometry* **32** (1969) 425–443; doi:10.2307/2786545.

[39] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature* **393** (1988) 440–442; doi:10.1038/30918.

[40] Y. Zhou and H. Liu, "Stability of periodic solutions for an SIS model with pulse vaccination", *Math. Comput. Modelling* **38** (2003) 299–308; doi:10.1016/S0895-7177(03)90088-4.