

# Infectious disease control using contact tracing in random and scale-free networks

Istvan Z. Kiss<sup>†</sup>, Darren M. Green and Rowland R. Kao

*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*

Contact tracing aims to identify and isolate individuals that have been in contact with infectious individuals. The efficacy of contact tracing and the hierarchy of traced nodes—nodes with higher degree traced first—is investigated and compared on random and scale-free (SF) networks with the same number of nodes  $N$  and average connection  $K$ . For values of the transmission rate larger than a threshold, the final epidemic size on SF networks is smaller than that on corresponding random networks. While in random networks new infectious and traced nodes from all classes have similar average degrees, in SF networks the average degree of nodes that are in more advanced stages of the disease is higher at any given time. On SF networks tracing removes possible sources of infection with high average degree. However a higher tracing effort is required to control the epidemic than on corresponding random networks due to the high initial velocity of spread towards the highly connected nodes. An increased latency period fails to significantly improve contact tracing efficacy. Contact tracing has a limited effect if the removal rate of susceptible nodes is relatively high, due to the fast local depletion of susceptible nodes.

**Keywords:** contact tracing; random networks; scale-free networks

## 1. INTRODUCTION

Recent development of the theory of complex networks has been rapid, as their applicability to a wide range of practical problems becomes recognized. These include complex systems such as the Internet, the World-Wide Web, ecological webs, biochemical networks (for a review see Albert & Barabási 2002), and in epidemiology social networks (Liljeros *et al.* 2001; Jones & Handcock 2003; Hufnagel *et al.* 2004) and livestock networks (Christley *et al.* 2005). The frequent occurrence of complex networks in nature raises the question of how infectious disease transmission and its control are affected by the underlying contact structure.

A network can be described by a combination of its nodes and links, which respectively represent individuals, population of individuals or organizations and the interactions amongst them. The degree of a node is its total number of links  $k$ . The theory of random networks, where the probability that a randomly chosen node has  $k$  links,  $P(k)$ , and follows a binomial distribution with mean  $K = \langle k \rangle$ , is well established (Erdős & Rényi 1959). Random networks have been and are used to model many real networks (Keeling 1999; Huerta & Tsimring 2002; Eames & Keeling 2003). While there are problems with interpreting

empirical data describing real networks (Jones & Handcock 2003; Stumpf *et al.* 2005), it has been recognized (Albert & Barabási 2002; Christley *et al.* 2005) that the degree distributions of many empirically measured networks do not follow a Binomial distribution (Poisson distribution in case of large networks), but can be approximated by a scale-free (SF) distribution  $P(k) \propto k^{-\gamma}$ , with  $2 \leq \gamma \leq 3$ . SF networks (Albert & Barabási 2002) have very different features when compared with standard random networks, and are characterized by high heterogeneity in node connectivity, with most of the nodes having a small number of links, while a relatively small number of nodes have a very large number of connections.

It is well known that an epidemic can spread on an infinitely large SF network, if the power law exponent of the distribution lies between 2 and 3 and the connections amongst nodes in the network are completely random, even for infinitesimally small transmission rates (Anderson & May 1991; Pastor-Satorras & Vespignani 2001), due to the diverging second moment of the nodes' degree distribution. However, May & Lloyd (2001) have shown that in finite size SF networks infections cannot spread for arbitrarily small transmission rates; nevertheless, the epidemic threshold in the transmission rate is much lower than in corresponding random networks.

Empirical studies have also shown the relevance of SF contact structures for epidemiological networks in the case of social contact networks (Liljeros *et al.* 2001; Jones & Handcock 2003; Hufnagel *et al.* 2004) and in the case of the cattle movement network within

<sup>†</sup>Author for correspondence (istvan.kiss@zoo.ox.ac.uk).

The electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2005.0079> or via <http://www.journals.royalsoc.ac.uk>.

Great Britain (Christley *et al.* 2005), even where determination of the exponent in the power law is problematic (Jones & Handcock 2003; Stumpf *et al.* 2005). Even though some diseases spread on networks with different characteristics such as small-world properties or any other non-trivial clustering or correlations, the role of high heterogeneity in the number of contacts of different individuals has long been recognized (Anderson & May 1991; Pastor-Satorras & Vespignani 2001) and deserves further consideration.

Contact tracing is a widely used epidemic control measure that aims to identify infected cases as early as possible (e.g. before clinical signs are observed) by following the contacts of individuals that are known to be infectious. While contact tracing was mostly successful for SARS (Lipsitch *et al.* 2003; Riley *et al.* 2003) and is useful for STDs (Clarke 1998; FitzGerald *et al.* 1998; Macke & Maher 1999), contact tracing failed to make a real impact in the recent 2001 foot-and-mouth disease epidemic (FMD) in the UK (Ferguson *et al.* 2001*a,b*; Keeling *et al.* 2001; Kao 2003), most likely due to inexperience in identifying dangerous contacts, and insufficient resources to remove them quickly (Haydon *et al.* 2004). Individuals that are identified by tracing can be isolated and quarantined with the possibility of recovery (e.g. SARS) or removed and culled (e.g. 2001 FMD epidemic in the UK). These different scenarios are all identical for the purpose of the model; however, they have very different implications for the traced individuals.

Real-world implementation of contact tracing may result in a high proportion of removed nodes actually being susceptible rather than infected. The proportion of susceptible nodes removed by tracing will depend on the disease (i.e. how easy it is to define what a truly dangerous contact is), the availability and quality of early diagnostic tests, and logistical constraints such as number of qualified personnel involved in contact tracing. Contact tracing where only infectious nodes are traced has been considered in the context of randomly distributed networks (Huerta & Tsimring 2002) and randomly distributed clustered networks (Eames & Keeling 2003). Recently Kiss *et al.* (2005) have considered the effect of removing susceptible nodes by tracing on random and clustered networks. Surprisingly, no study has yet directly compared disease control on SF and random networks for given epidemiological parameters. This is done here, first for a simple disease model with no control, and then allowing for the identification and removal of nodes that are at risk of becoming infected—i.e. ‘contact tracing’—is considered.

Targeting control at those nodes that are most responsible for disease spread is an attractive way of reducing the effort required to control it. The importance of highly connected ‘superspreaders’ has been discussed in detail for more abstract models (Anderson & May 1991), and more recently the role of highly connected nodes for SF networks has been noted (Albert *et al.* 2000; Barthélemy *et al.* 2004). Optimal preventive control measures have received the most attention thus far, in particular related to vaccination.

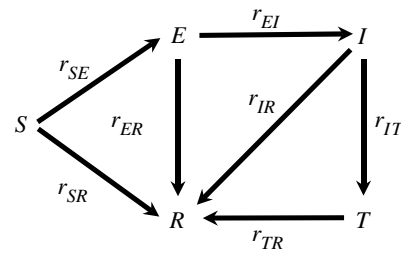


Figure 1. Possible transitions among the five different classes and the corresponding rates of transitions.

The most widely discussed are the targeting for immunization of random nodes, and immunization of random acquaintances of random nodes (Madar *et al.* 2004). However, if an epidemic has already become established in a population and the turnover time (i.e. the time interval between consecutive generations of infectious individuals) of the disease is relatively small (such as SARS, FMD, etc.), then the main control measure that can be used is disease contact tracing. This targets the nodes thought to be at risk of becoming infected with the aim of reducing the infectious period, preferably to zero.

## 2. MODEL

### 2.1. Disease transmission model and individual-based stochastic simulation

Each node in the network is classified according to one of five different states (classes): susceptible nodes ( $S$ ); exposed nodes ( $E$ ), i.e. incubating the disease without being infectious; infectious nodes ( $I$ ); triggering nodes ( $T$ ), which are identified as being infectious, immediately isolated, and initiating tracing; and finally, removed nodes ( $R$ ), which are no longer infectious and do not initiate tracing. The possible transition between states are depicted in figure 1 and presented in detail below.

- *Infection*  $S \rightarrow E$ . The epidemic is seeded with one or more infected nodes. Thereafter, infection progresses as a simple contact process: the probability of a secondary infection depends on the state of the neighbouring nodes.  $r_{SE}$  is the infection rate which represents the probability per unit time of a susceptible node acquiring infection upon contact with an infectious node. The exposed state represents the delay between becoming infected and being infectious.
- *Becoming infectious*  $E \rightarrow I$ . This is the natural transition of the exposed nodes to the infectious class after the latency period has elapsed, with rate  $r_{EI}$ . This is a network-independent process.
- *Direct identification of infectious nodes*  $I \rightarrow T$ . Disease detected at an infectious node will trigger tracing.  $r_{IT}$  is the transition rate of the infectious nodes to class ( $T$ ), and is a network-independent process.
- *Removal of class ( $T$ ) nodes*  $T \rightarrow R$ . This is the network-independent removal of the class ( $T$ ) nodes with a given rate  $r_{TR}$ . We assume that no

further tracing occurs from a removed node. While this is not strictly true, the assumption simplifies the system, and the effect of continued tracing can be approximated by increasing the tracing rate from class ( $T$ ) nodes. Further, late tracing is likely to be less important for diseases with short turnover time, as many infected nodes would likely become positively diagnosed *via*  $I \rightarrow T$ .

- *Tracing ‘errors’*  $S \rightarrow R$ . Contact tracing often identifies some contacts that were not truly infectious (e.g. through human error, logistical constraints compromising testing accuracy, or a lack of rapid and conclusive diagnosis tests). It is assumed that a node identified through contact tracing cannot initiate further tracing; i.e. ‘secondary’ tracing is suppressed. Therefore, a direct contact between a class ( $T$ ) node and a susceptible node, detected through tracing, results in the direct transition of the susceptible node into the removed class ( $R$ ). The rate of this process is  $r_{SR}$ .
- *Tracing and removal of exposed nodes*  $E \rightarrow R$ . Exposed nodes in direct contact with class ( $T$ ) nodes may be identified as potentially infectious and are removed with rate  $r_{ER}$ .
- *Tracing and removal of infectious nodes*  $I \rightarrow R$ . As with tracing of susceptible and exposed nodes, tracing of infectious nodes does not trigger secondary tracing. This process happens at a rate  $r_{IR}$ .

While in principle traced nodes might trigger further tracing, in this model we assume that secondary tracing (i. e. tracing from traced  $E$  and  $I$  nodes) does not occur. The absence of the secondary tracing is an appropriate assumption in various situations, such as for livestock diseases where disease control policies only allow for the tracing and removal of those premises where contact with an infectious premises can be proven. This is a difficult task even in the first step contact tracing, therefore further contact tracing is very rare. Another impediment is the difficulty in determining the status of the traced premises due to the absence of a quick and reliable diagnostic test for some diseases (e.g. FMD). Multiple-stage contact tracing is usually carried out where traced individuals are isolated and can be monitored for signs of infection. This is not the case for livestock diseases where traced premises are usually culled. Müller *et al.* (2000) also concludes that tracing a very small number of steps from the index case may already practically maximize the effect that can be reached by tracing. In the event of logistical constraints, it is likely that effort will be concentrated on primary tracing, and so the question of the conditions under which primary tracing alone is useful, is an important one.

Computer-generated random and SF networks are used to simulate disease propagation based on a synchronous updating scheme. The state of a node at time  $t_0 + \Delta t$  is based solely on the state of its neighbours at time  $t_0$  and the rates defining the transitions between the states.

## 2.2. Network structure and parameter values

Random networks are generated by considering all  $N(N-1)/2$  possible links, and, for each link, comparing random numbers to the threshold probability  $p$ . Any link whose random number is greater than  $p$  is discarded. The expected value of the average number of connections per node  $K$  is then  $p(N-1)$ .

SF networks are generated according to the method described by Newman *et al.* (2001). The degree distribution of the SF graph is given by  $P(k) = Ck^{-\gamma} \exp(-k/L)$ , where  $C$  is the normalizing constant,  $\gamma$  is the power law exponent and  $L$  is the natural cut-off present in any finite-size natural system. The generation of such a network is done as follows: for each node a random integer  $k \geq 3$  is generated, as a potential degree, from a distribution proportional to  $\exp(-k/L)$  using the transformation  $k = \text{ceil}(-L \ln(1-r))$ , where  $r$  is a uniformly distributed random number in the range  $0 < r < 1$  and  $\text{ceil}()$  is a function that returns the closest integer number bigger than its argument. Only nodes with  $k \geq 3$  are chosen, to support a realistic epidemic where  $K$  must exceed the number of new infections in the first generation. The value of  $k$  is accepted as a node degree with a probability  $k^{-\gamma}$ . If  $k$  is not accepted then it is discarded and a new  $k$  is generated. The procedure is repeated until a value for  $k$  is accepted for every node in the network and the sum of the degree of all the nodes is even. References to each node are placed in a set such that the number of references to each node is equal to the node degree. Thereafter, two members are chosen at random from the set and a link is formed if these nodes are not already connected. Self-loops are not considered as possible links. This operation is repeated until all the members of the set are exhausted. The even number of members ensures that after repeated link forming operations all the nodes will have a number of links equal to their degree. All the random and SF networks generated in this paper have  $N=2000$  nodes and the average number of connection per node  $K=6$ . All the SF networks are generated using  $\gamma=2.5$ ,  $L=100$ .

If  $\text{Inf\_P}$  is the average time spent by nodes in the infectious class in the absence of network related tracing then  $r_{IT} = 1/\text{Inf\_P}$ . Similarly, if  $\text{Lat\_P}$  and  $\text{Tr\_P}$  are the average time periods in days spent by nodes in the exposed and traced classes, respectively, then  $r_{EI} = 1/\text{Lat\_P}$  and  $r_{TR} = 1/\text{Tr\_P}$ . Following from Kiss *et al.* (2005), throughout the paper  $\text{Inf\_P}=3.5$  and  $\text{Tr\_P}=2.0$  are used. In the numerical simulations the transmission rate, the network dependent tracing rates and  $\text{Lat\_P}$  are varied in order to investigate how contact tracing impacts on disease spread. These parameter values are similar to the latency, infectious and tracing periods for diseases such as SARS (Lipsitch *et al.* 2003; Riley *et al.* 2003) and FMD (Ferguson *et al.* 2001a,b; Keeling *et al.* 2001; Kao 2003), and are also similar to the parameters used in other epidemiological network studies (Huerta & Tsimring 2002; Eames & Keeling 2003; Green *et al.* in press).

### 3. RESULTS

#### 3.1. Epidemic threshold in random and scale-free graphs

The time evolution of the proportion of infectious nodes is given for random networks in figure 2a and SF networks in figure 2b for different values of  $r_{SE}$  and no tracing. For both cases  $\text{Lat}_P=3.5$ . The results are averaged over 50 different network realizations and 50 different epidemics on each individual network realization. Each individual epidemic on each network realization is seeded with 10 randomly placed infectious nodes. An increase in any of the two numbers presented above or a finer resolution in time than  $\Delta t=0.02$  produced effectively the same results. As expected the threshold value of  $r_{SE}$  for disease persistence is higher on random networks than on corresponding SF networks. In the case of SF networks, due to the diverging second moment ( $\langle k^2 \rangle$ ) of the nodes' degree distribution the bigger the network size  $N$ , the smaller the threshold for  $r_{SE}$  is. In the limit of infinite-size SF networks, an infinitesimally small but non-zero  $r_{SE}$  will result in a well-established epidemic in the population (Anderson & May 1991; Pastor-Satorras & Vespignani 2001). As expected, the initial spread of the disease on SF networks, driven by the highly connected nodes, is more rapid than for random networks.

In figure 3 the final epidemic size, representing the final proportion of the nodes that became infectious during the epidemic, is given as a function of  $r_{SE}$  for random and SF networks with no tracing applied. For smaller values of  $r_{SE}$ , the final epidemic size for SF networks is considerably higher than that for random networks. However, as  $r_{SE}$  increases, there is a critical value of  $r_{SE}$  where the final epidemic sizes are the same for both types. From there on, final epidemic size for random networks approaches its asymptote (total population size) more rapidly than for SF networks. As  $r_{SE}$  increases, resulting in a higher total epidemic size, the disease depletes the population of highly connected nodes, and must propagate to the poorly connected nodes, which are less accessible. Since the random networks have a more homogeneous degree distribution than SF networks, this depletion occurs more slowly, so the final epidemic size is larger than on equivalent SF networks for large  $r_{SE}$ .

#### 3.2. The effect of contact tracing in random and scale-free networks

The role of contact tracing is considered on different network architectures for a wide combination of  $r_{SR}$ ,  $r_{ER}$  and  $r_{IR}$  tracing rates. We consider only equal tracing rates of exposed and infectious nodes:  $r_{ER}=r_{IR}$ . The primary aim of tracing is the removal of as many infected (exposed or infectious) nodes as possible, since only these can transmit disease. However, as mentioned above, the removal of susceptible nodes is unavoidable, and therefore the interaction between the tracing rate of susceptible nodes ( $r_{SR}$ ) and the tracing rates of infected nodes ( $r_{ER}=r_{IR}$ ) is investigated to establish the overall effect of tracing. Only the  $r_{ER}=r_{IR} \geq r_{SR}$  case is investigated because is unlikely that susceptible

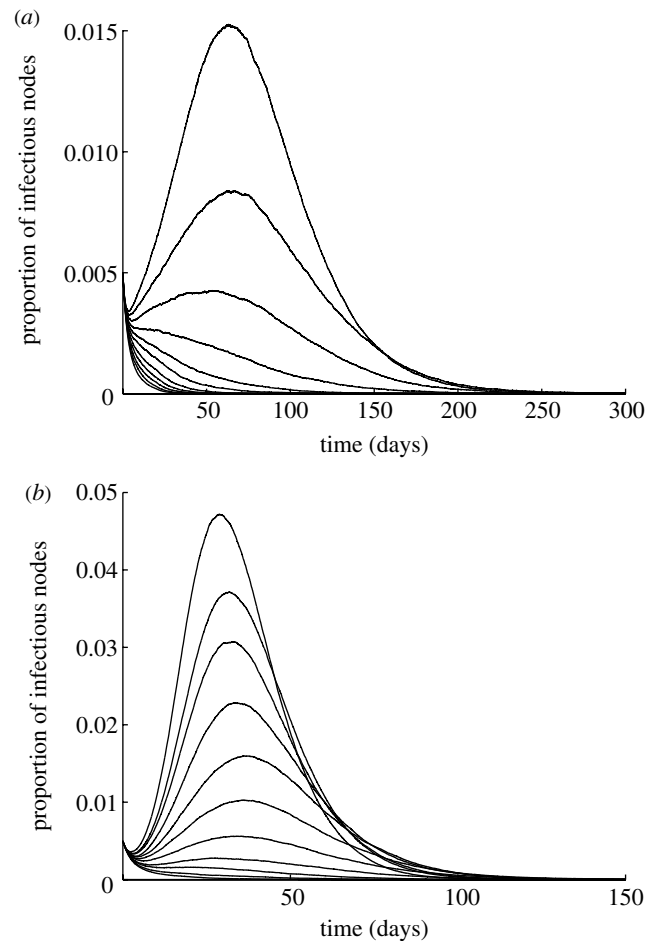


Figure 2. Time evolution of the proportion of infectious nodes for random (a) and scale-free (SF), (b) networks in the absence of tracing with  $r_{SE}=0.015, 0.0215, \dots, 0.08$  and  $\text{Lat}_P=3.5$ .

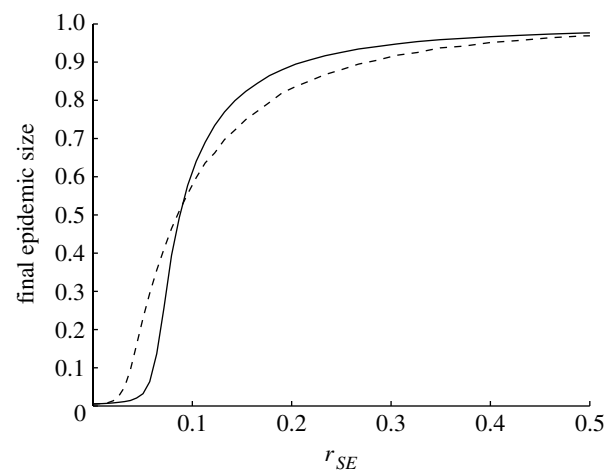


Figure 3. Final epidemic size representing the number of nodes that became infectious during the whole epidemic, plotted as a function of  $r_{SE}$  in the absence of tracing. Continuous line corresponds to random networks and dashed line to SF networks. The latency period is  $\text{Lat}_P=3.5$ .

nodes are traced at a higher rate than exposed and infectious nodes.

In figure 4 the proportion of the total number of infected nodes removed by tracing is given as a function

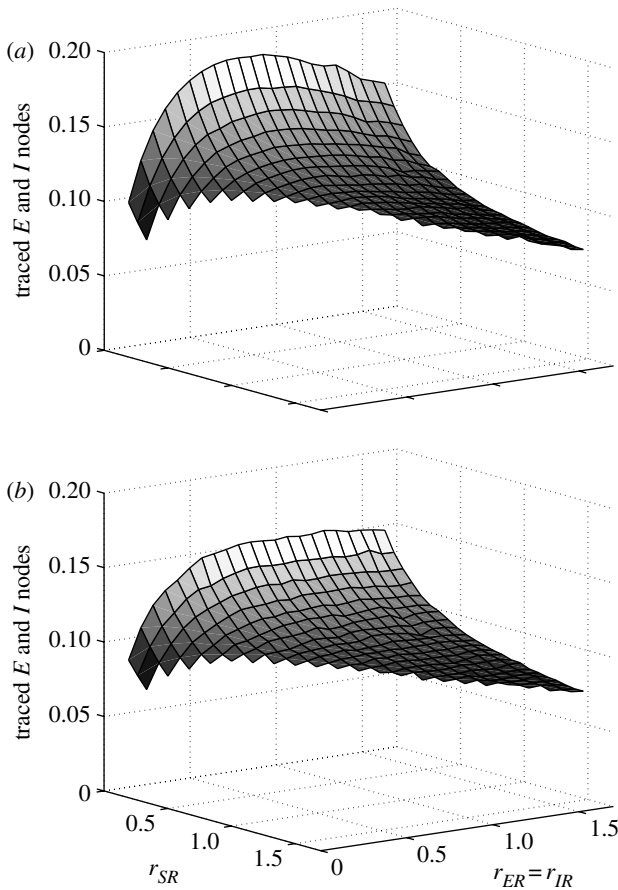


Figure 4. Proportion of infected nodes removed by tracing for random (a) and SF (b) networks for  $r_{SE}=0.15$  and  $\text{Lat\_P}=3.5$ .

of the different tracing rate combinations for random (figure 4a) and SF (figure 4b) networks with  $r_{SE}=0.15$  and  $\text{Lat\_P}=3.5$ . The value of  $r_{SE}=0.15$  was chosen, as at this value, the transmission potential  $\rho_0 = r_{SE} K \text{Inf\_P}$  as defined by May & Lloyd (2001) is similar to the values that were estimated from the recent SARS epidemic (Lipsitch *et al.* 2003) and the 2001 UK FMD epidemic (Ferguson *et al.* 2001a).

Consider the case when  $r_{SR}$  is fixed and  $r_{ER}(=r_{IR})$  is varied. If the epidemic is out of control, an increase in  $r_{ER}(=r_{IR})$  results in more infected nodes being removed. While the current reproduction ratio of the epidemic remains above one, the epidemic incidence will on average continue to increase, and the trend of increasing numbers of removed infected nodes can be expected to continue, as long as the total number of susceptible nodes is not depleted. However, should higher rates of tracing  $r_{ER}(=r_{IR})$  be sufficient to reduce the reproduction rate below one, the proportion of infected nodes removed by tracing will start to decrease; i.e. tracing will be considered to have effectively controlled the epidemic. Thus, although  $r_{ER}(=r_{IR})$  is high, the proportion of infected nodes removed by tracing is small since the production of new infected nodes is reduced. Therefore, for a fixed value of  $r_{SR}$ , the maximum in the surface for a specific value of  $r_{ER}(=r_{IR})$  represents the threshold for the minimum level of tracing capable of controlling the epidemic.

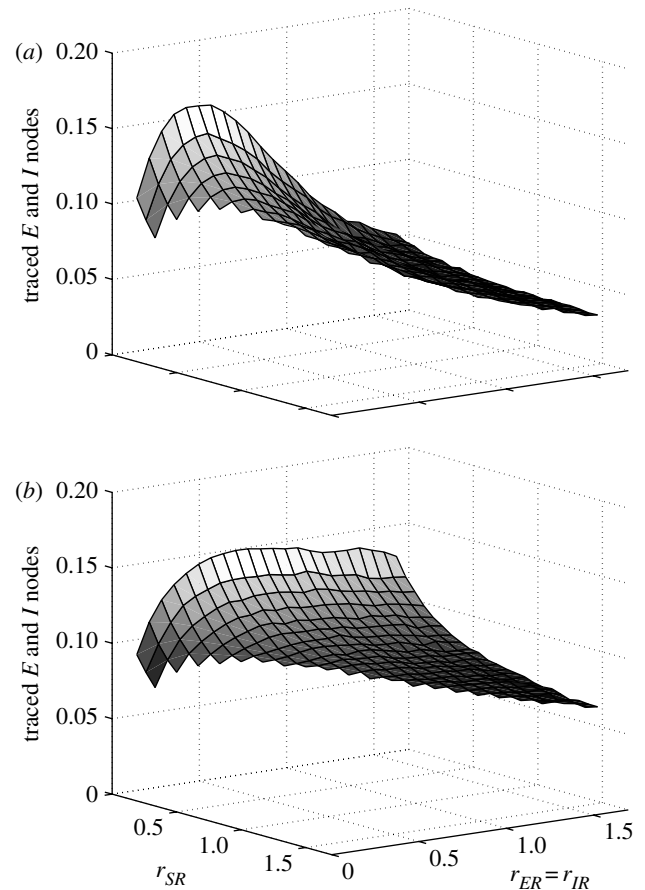


Figure 5. Proportion of infected nodes removed by tracing for random (a) and SF (b) networks for  $r_{SE}=0.15$  and  $\text{Lat\_P}=10.0$ .

According to figure 4 contact tracing on random networks is effective for values of  $r_{ER}(=r_{IR})$  that are lower than those needed on the corresponding SF networks, if the same constant  $r_{SR}$  is considered for both networks. This is shown by the decline in the proportion of infected nodes removed by tracing at lower tracing rates  $r_{ER}(=r_{IR})$  in random networks, as compared to SF networks. This observation suggests that on SF networks a tracing effort higher than that on similar random networks is required to control and stop an epidemic spreading. Individually, the proportions of exposed and infectious nodes exhibit the same qualitative behaviour as their sum.

An important role in the control of an epidemic is played by the length of the latency period  $\text{Lat\_P}$  (Fraser *et al.* 2004; Kiss *et al.* 2005). Intuitively, a longer latency period favours contact tracing by allowing a longer time for the identification of possible new infectious cases, especially for nodes that are exposed and spend a considerable period in the latent state without being infectious. However, a short latency period will make contact tracing difficult due to the quick turnover time of the newly infected generation of nodes. Figure 5 is the same as figure 4 but with an increased latency period from  $\text{Lat\_P}=3.5$  to  $\text{Lat\_P}=10.0$ . In this case, on random networks contact tracing performs considerably better than in the  $\text{Lat\_P}=3.5$  case. For SF networks however, only a comparatively limited improvement is noticeable.

Contact tracing is more effective on random networks and manages to control the epidemic with smaller tracing rates  $r_{ER}(=r_{IR})$ . The proportion of nodes that become infectious by the end of the epidemic and the proportion of nodes removed by the end of the epidemic ( $R_\infty$ ) are also plotted in figures S1, S2, S3 and S4 (see electronic supplementary material) for the same parameter values as figures 4 and 5.

With higher tracing rates, both the proportion of nodes that become infectious by the end of the epidemic and the proportion of nodes removed by the end of the epidemic decrease monotonically. However, the presence of maxima in the proportion of infected nodes removed by tracing (figures 4 and 5) illustrates a change in the epidemic behaviour for higher tracing rates; here the number of infected nodes that are traced decreases despite the increased tracing rates. This suggests that tracing is becoming effective on a local neighbourhood level at higher tracing rates (Kiss *et al.* 2005).

Barthélemy *et al.* (2004) have shown that the initial exponential growth time-scale  $\tau$  of epidemics is inversely proportional to the network degree fluctuations,  $\langle k^2 \rangle / \langle k \rangle$ . In networks with a Poisson distribution (such as large random networks)  $\langle k^2 \rangle / \langle k \rangle = \langle k \rangle + 1$ , however in SF networks and in general in networks with highly heterogeneous connectivity,  $\langle k^2 \rangle / \langle k \rangle \gg \langle k \rangle$ . This implies an extremely small time-scale  $\tau$  for the outbreak and a very rapid spread of the epidemic in networks with highly heterogeneous connectivity. Barthélemy *et al.* (2004) have also shown that the disease spread follows a precise hierarchical order, with the highly connected nodes becoming infected first, and the epidemic thereafter cascading towards groups of nodes with smaller degree. This can be seen in figure 6*b*, where the average degree of new infectious nodes is plotted in the presence of tracing.

To better understand the effect and the mechanism of tracing, the average degree of new infectious nodes, and the average degree of susceptible and infected nodes removed by tracing are plotted in figure 6. We note that in the absence of tracing the change in the average degree of new infectious nodes follows a similar pattern. In random networks (figure 6*a*) the average degree of all the different categories of nodes are similar throughout the epidemic time course. All types start at  $\langle k^2 \rangle / \langle k \rangle = \langle k \rangle + 1 = K + 1$  which is the average degree of nodes that can be reached from a randomly selected node in any network. The average degree of newly infected nodes and nodes traced from a given ( $T$ ) class node reflect only this network property (as in the ‘acquaintance sampling’ described by Madar *et al.* (2004)) and thus they are of the same degree. This is seen in figure 6*a* for random networks, and figure 6*b* for SF networks. However, the average degrees of both exposed and infectious nodes removed by tracing are considerably higher than the former categories of nodes. This is a finite size effect; nodes of high degree tend to be infected first and as these are depleted the average degree decreases. Thus the ‘age’ of traced nodes must be considered. The average duration that nodes spend in state  $E$  and state  $I$ , considered from when the node became infected in the absence of network related tracing, are  $1/r_{EI}$  and  $(1/r_{EI} + 1/r_{IT})$  respectively.

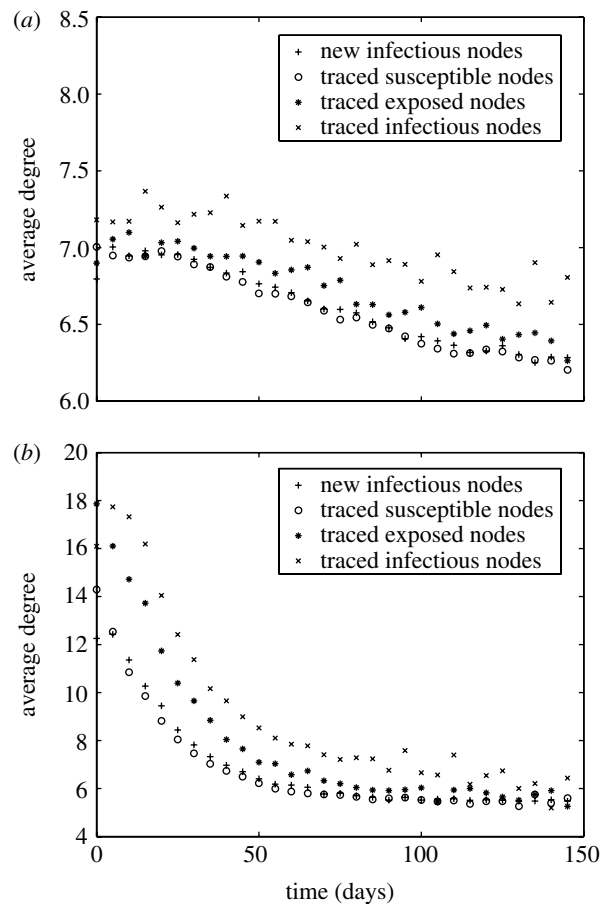


Figure 6. Average degree of new infectious nodes, and average degree of susceptible and infected nodes removed by tracing *versus* time for random (a) and SF (b) networks for  $r_{SE}=0.15$ ,  $\text{Lat}_P=10.0$  and  $r_{SR}=r_{ER}=r_{IR}=0.5$ .

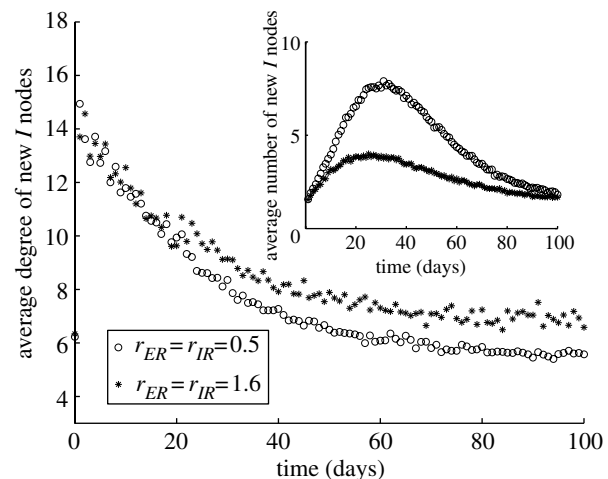


Figure 7. The daily average degree and number (inset) of new infectious nodes *versus* time for different levels of tracing ( $r_{ER}=r_{IR}=0.5$  and  $1.6$ ). Results obtained using SF networks for  $r_{SE}=0.15$ ,  $\text{Lat}_P=10.0$  and  $r_{SR}=0.5$ .

Traced nodes arising from these classes reflect the average degree of nodes infected at an earlier time (figure 6).

To further investigate the mechanism of tracing, in figure 7 the daily average degree and number (inset) of new infectious nodes is plotted for a constant tracing rate  $r_{SR}$  of the susceptible nodes and different levels of

tracing of infected nodes  $r_{ER}(=r_{IR})$  for SF networks. In the case without control, the disease preferentially spreads towards the highly connected nodes. Due to depletion of highly connected nodes, proportionately more nodes with lower connectivity are infected towards the end of the epidemic. When effective control is applied, depletion of highly connected nodes occurs more slowly, and therefore the average degree of new infectious nodes remains high. Thus, the average degree of the new infectious nodes at higher levels of tracing will be more evenly distributed over the duration of the epidemic, in contrast to the highly skewed distribution when disease is spreading without control. The inset of figure 7 shows that the daily average number of new infectious nodes falls sharply as the  $r_{ER}(=r_{IR})$  tracing rate is increased. Therefore, the higher daily average degree of new infectious nodes that corresponds to higher levels of tracing  $r_{ER}(=r_{IR})$  later is indicative of superior but still ineffective control.

#### 4. DISCUSSION

While epidemic spread on SF networks has recently been the subject of much interest, to our knowledge there has been no quantitative comparison of the differences in disease control efficacy on SF and random networks. Our analysis confirms that in random networks, epidemic spread is much slower, and the average degree of new infectious nodes is close to the average degree  $K=\langle k \rangle$  (figure 6a). The relatively slow time-scale of the disease, and the virtually degree-independent spread of the disease help contact tracing to follow the disease closely and to control it if the tracing rate is high enough.

In SF networks the disease turnover time is fast and the disease preferentially spreads to nodes with high degree (figure 6b); however, highly connected nodes are also typically traced at a later disease stage (equivalently, the average degree of exposed and traced nodes is delayed with respect to the degree of traced and infectious nodes). This is directly connected to the latency period, since infectious nodes traced at a given time in the epidemic represent a set of nodes that were infected on average a time  $\text{Lat}_P$  earlier in the epidemic, and therefore are of higher average degree than the traced and exposed nodes.

Unlike for random networks, even a substantial increase in the latency period has only a limited effect on the efficacy of contact tracing on SF networks. Examination of figure 6 suggests that while acquaintance sampling (Madar *et al.* 2004) may be an effective way of targeting proactive vaccination on a SF network, at least for the parameter regimes studied here, it is not an effective strategy for reactive control. Contact tracing never catches up sufficiently to eliminate the highly connected nodes before they become infectious, and only ‘intelligent’ tracing (for example through prior knowledge of who may be most highly connected) would be effective.

Figures S3 and S4 (see electronic supplementary material) show, for the same networks and parameter values as figures 4 and 5, that  $R_\infty$  is insensitive relative to changes in  $r_{SR}$ . Therefore unless tracing of exposed

and infected nodes happens quickly and accurately the loss encountered at the population level is virtually constant with regards to changes in  $r_{SR}$ . Disease contact tracing can make a real difference on both random and SF networks when the tracing rate of susceptible nodes is low. Otherwise, the depletion of susceptible nodes is the dominant effect and the disease burns itself out due to the lack of new targets.

An epidemic controlled by contact tracing can end either due to the early and precise identification of most infected nodes, or due to the depletion of available susceptible nodes where the disease could have propagated. A control policy that results in removal of many susceptibles can be the result of many factors, including human error, time constraints in determining whether a node is truly infected, an insufficiently specific diagnosis test, or deliberate policy. There may be significant economic or moral cost associated with removal of susceptible nodes; the main advantage in this case is a shorter epidemic duration. Ideal control measures; however, should stop an epidemic early or while there are still many unaffected susceptible nodes. Should contacts between nodes be highly clustered, the removal of susceptible nodes can reduce the local susceptible neighbourhood surrounding infectious nodes (Eames & Keeling 2003; Kiss *et al.* 2005), rendering it an effective means of disease control. This is often acceptable, for example when removal by tracing actually means quarantine or isolation (such as SARS), or if the consequences of transmission are severe, but is much more debatable when traced nodes are destroyed, especially when there is some controversy over how seriously the disease problem is viewed (such as 2001 FMD epidemic in UK; Haydon *et al.* 2004). Robust estimations of the clustering coefficient are critical for evaluation of any such policy.

A major role in infectious disease spread is played by the underlying contact structure. However, in most cases, the exact contact structure is difficult to determine. It is, therefore, imperative to understand how to best use the available data about a network on which disease can spread in order to better prevent or control epidemics. By analysing disease transmission on different network topologies, one can hope to identify trademark signatures of different disease models spreading on different network architectures. Such results might help solve the inverse problem of what network or disease parameters can be recovered from data collected during an epidemic, which can be critical in deciding amongst different control policies.

R.R.K. and I.Z.K. are funded by the Wellcome Trust. D.M.G. is funded by DEFRA.

#### REFERENCES

- Albert, R. & Barabási, A. L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)
- Albert, R., Jeong, H. & Barabási, A. L. 2000 Error and attack tolerance of complex networks. *Nature* **406**, 378–382. (doi:10.1038/35019019)

- Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press.
- Barthélemy, M., Barrat, A., Pastoras-Satorras, R. & Vespignani, A. 2004 Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178701. (doi:10.1103/PhysRevLett.92.178701)
- Clarke, J. 1998 Contact tracing for Chlamydia: data on effectiveness. *Int. J. STD AIDS* **9**, 187–191. (doi:10.1258/0956462981921945)
- Christley, R. M., Robinson, S. E., Lysons, R. & French, N. P. 2005 Network analysis of cattle movement in Great Britain. In *Proc. Soc. for Veterinary Epidemiology and Preventive Medicine* (ed. D. J. Mellor, A. M. Russell & J. L. N. Wood), Nairn, Scotland, 30 March–1 April, pp. 234–244.
- Eames, K. T. D. & Keeling, M. J. 2003 Contact tracing and disease control. *Proc. R. Soc. B* **270**, 2565–2571. (doi:10.1098/rspb.2003.2554)
- Erdős, P. & Rényi, A. 1959 On random graphs. *Publ. Math.* **6**, 290–297.
- FitzGerald, M. R., Thirlby, D. & Bedford, C. A. 1998 The outcome of contact tracing for gonorrhoea in the United Kingdom. *Int. J. STD AIDS* **9**, 657–660. (doi:10.1258/0956462981921305)
- Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001a The foot-and mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160. (doi:10.1126/science.1061020)
- Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001b Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542–548. (doi:10.1038/35097116)
- Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. 2004 Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151. (doi:10.1073/pnas.0307506101)
- Green, D. M., Kiss, I. Z., & Kao, R. R. In press. Parameterisation of individual-based models: recipes for translation from deterministic mean-field models. *J. Theor. Biol.*
- Haydon, D. T., Kao, R. R. & Kitching, R. P. 2004 Opinion: the UK foot-and-mouth disease outbreak—the aftermath. *Nat. Rev. Microbiol.* **2**, 675–681. (doi:10.1038/nrmicro960)
- Huerta, R. & Tsimring, L. S. 2002 Contact tracing and epidemics control in social networks. *Phys. Rev. E* **66**, 056 115. (doi:10.1103/PhysRevE.66.056115)
- Hufnagel, L., Brockmann, D. & Geisel, T. 2004 Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15 124–15 129. (doi:10.1073/pnas.0308344101)
- Jones, H. J. & Handcock, M. S. 2003 An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. R. Soc. B* **270**, 1123–1128. (doi:10.1098/rspb.2003.2369.)
- Kao, R. R. 2003 The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease. *Proc. R. Soc. B* **270**, 2557–2564. (doi:10.1098/rspb.2003.2546)
- Keeling, M. J. 1999 The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. B* **266**, 859–867. (doi:10.1098/rspb.1999.0716)
- Keeling, M. J. *et al.* 2001 Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817. (doi:10.1126/science.1065973)
- Kiss, I. Z., Green, D. M. & Kao, R. R. 2005 Disease contact tracing in random and clustered networks. *Proc. R. Soc. B* **272**, 1407–1414. (doi:10.1098/rspb.2005.3092)
- Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Aberg, Y. 2001 The web of human sexual networks. *Nature* **411**, 907–908. (doi:10.1038/35082140)
- Lipsitch, M. *et al.* 2003 Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970. (doi:10.1126/science.1086616)
- Macke, B. A. & Maher, J. E. 1999 Partner notification in the United States: an evidence-based review. *Am. J. Prev. Med.* **17**, 230–242. (doi:10.1016/S0749-3797(99)00076-8)
- May, R. M. & Lloyd, A. L. 2001 Infection dynamics on scale-free networks. *Phys. Rev. E* **64**, 066112. (doi:10.1103/PhysRevE.64.066112)
- Müller, J., Kretzschmar, M. & Dietz, K. 2000 Contact tracing in stochastic and deterministic epidemic models. *Math. Biosci.* **164**, 39–64.
- Madar, N., Kalisky, T., Cohen, R., ben-Avraham, D. & Havlin, S. 2004 Immunization and epidemic dynamics in complex networks. *Eur. Phys. J. B* **38**, 269–276.
- Newman, M. E. J., Strogatz, S. H. & Watts, D. J. 2001 Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E* **64**, 026118. (doi:10.1103/PhysRevE.64.026118)
- Pastor-Satorras, R. & Vespignani, A. 2001 Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)
- Riley, S. *et al.* 2003 Transmission dynamics of the aetiological agent of severe acute respiratory syndrome SARS in Hong Kong: the impact of public health interventions. *Science* **300**, 1961–1966. (doi:10.1126/science.1086478)
- Stumpf, M. P. H., Wiuf, C. & May, R. M. 2005 Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc. Natl Acad. Sci. USA* **102**, 4221–4224. (doi:10.1073/pnas.0501179102)