

Network Effects in Epidemiology

Dr. Anthony H. Dekker

Defence Science and Technology Organisation (DSTO)
Canberra, Australia
dekker@ACM.org

Abstract. Preparing for the spread of infectious diseases is an important government function. The great influenza pandemic of 1918 killed about 100 million people, and in more recent times, SARS and avian influenza have affected several countries. Agent-based models can be used to predict the spread of new diseases within a population. However, the behaviour of such models is dependent on the topology of contacts between people. In this paper, we use rewiring and simulation techniques to investigate the spread of disease through a network, and highlight the network attributes which influence the spread of disease. We pay particular attention to three key disease characteristics: the peak number of infected individuals, the time to reach this peak infection, and the total number of individuals affected by the disease. Analysis of the results leads to a concept of “effective degree” for a network, which depends on the number of links in the network as well as the network “shape.” The three key characteristics are all strongly influenced by the effective degree.

1. INTRODUCTION

Pandemics of infectious disease are a serious threat to society. The great influenza pandemic of 1918 killed about 100 million people [1], and in more recent times, SARS and avian influenza have affected several countries. There is also the threat of pandemic resulting from the release of biological weapons by terrorists.

Infectious diseases spread through the human social network, and network effects are significant in influencing the spread of disease. In this paper, we use a simple agent-based simulation model together with network rewiring techniques [4] to explore these network effects.

2. THE SIR MODEL

The spread of disease through a population can be modelled mathematically as a set of differential equations. Although several such mathematical models exist, the SIR equations [2] are most often used:

$$\frac{dS}{dt} = -\alpha SI \quad (1)$$

$$\frac{dI}{dt} = \alpha SI - \beta I \quad (2)$$

$$\frac{dR}{dt} = \beta I \quad (3)$$

where S is the number of individuals in the population susceptible to the disease, I is the number currently infected, R is the number recovered (and resistant), and α and β are parameters. The graph in Figure 1 shows the change in the variables with time (from an initial state with $I = 1$, $S = 399$, and $R = 0$) for a particular choice of parameters α and β .

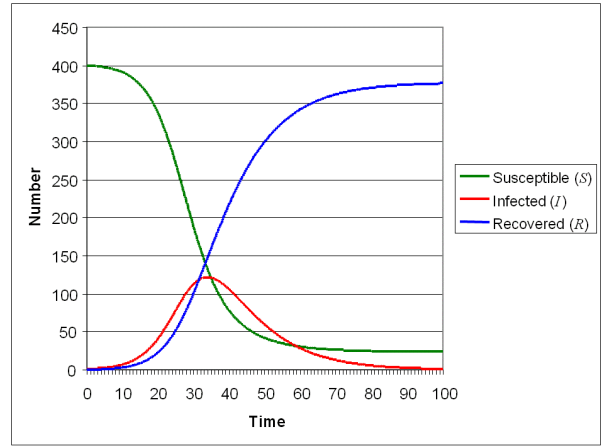


Figure 1: Example plot for the SIR equations (for $I = 1$ and $S = 399$ initially, with $\alpha = 0.00075$, $\beta = 0.1$)

As in real disease outbreaks, there is a “peak” of the disease, where the number of infected individuals I reaches a maximum, followed by a decline. The time to reach this peak is of interest to makers of health policy, since it determines the time available to respond to early signs of a pandemic. The peak number of infected individuals is also important, determining the scale of medical facilities required. Finally, the cost of the disease – both in dollars and in terms of human suffering – is determined by the total number of people infected, i.e. the final value of R . In a pandemic, this will be a significant fraction of the whole population.

A limitation of the SIR equations is that they ignore effects due to the network topology of contacts between people. We explore these effects using a very simple agent-based simulation model.

3. AGENT SIMULATION

For the agent-based model, it is convenient to have slightly different parameters. One of these is the duration δ of a person being infected (where $\beta \approx 1/\delta$). For our experiment, values of 1, 3, 10, 30, and 100 time units were used.

The parameter α is the probability of an individual infecting another linked individual during a single time unit (values of α are substantially higher than for the SIR equations, since an individual has relatively few contacts in general). However, rather than varying α directly, we found it convenient to vary $\gamma = \alpha\delta$, using the values 0.1, 0.3, 0.5, 0.7, 0.85, and 1.0. The parameter γ is approximately equal to the probability of infecting another linked individual over the duration of infection, and is largely independent of the choice of δ .

We combine this simple agent-based model with the Kawachi process for network rewiring [4],[5], which extends the rewiring process due to Watts [6], rearranging the links while keeping their number constant. We begin with a 4×100 rectangular grid, connecting the long edges to form a narrow tube of length 100, then connecting the ends to form a Torus, and finally adding diagonals across the grid squares to bring the degree of each node up to 6.

The Kawachi rewiring process, as we have implemented it, is controlled by a parameter p between 0 and 5 and takes place in a number of phases (10 phases for our experiment). In each phase, network links $x-y$ are replaced with probability $p/10$ by links $x-z$, where the person (node) x has a higher degree than y , and the node z is chosen to be an isolated node from a previous rewiring (if there are any), or chosen randomly (with probability proportional to the degree of nodes plus one) otherwise. For small values of p , this is equivalent to the Watts rewiring process, while for large values of p , the high-degree bias produces networks with Scale-Free characteristics [4],[5].

Figure 2 illustrates the impact of Kawachi rewiring. For $0.02 \leq p \leq 0.1$, the average network distance D between nodes has dropped rapidly from 25.6 down to about 6, giving Small-World networks. These networks, though having a low average distance, retain much of the original network structure, reflected in a clustering coefficient C close to the initial value of 0.4 [4],[5].

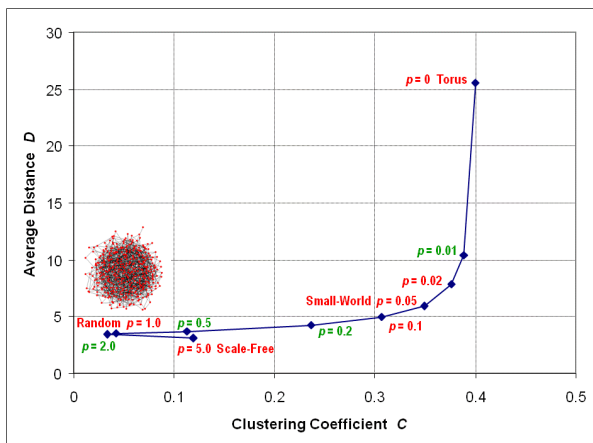


Figure 2: The Kawachi Process

Higher values of the Kawachi rewiring parameter p progressively destroy more of the original network structure, until at $p = 1$, Random networks are produced, with $D = 3.5$ and $C = 0.03$. These networks have links distributed essentially at random between nodes. One such network is shown on the left of Figure 2.

Further increases in p (up to 5) produce Scale-Free networks, which have a power-law distribution of node degrees, resulting in a small number of highly-connected “hub” nodes. Scale-Free networks have a slightly lower average distance D than Random networks, but a higher clustering coefficient C . The Kawachi process therefore transitions through four different kinds of network using a single parameter, which is convenient for experimental purposes. In our simulations, we generate 20 different networks for each value of p .

Figure 3 shows an example simulation snapshot with a Random network and $\gamma = 0.5$, $\delta = 10$. The peak of 172 infected individuals (coloured red, orange, and yellow for various stages of the infection duration) has just occurred at time 29. Now, at time 34, there are 136 infected individuals. So far, 182 individuals have recovered (blue), and 82 are still susceptible (green). For this network and combination of parameters, the infection will stop with about 350 recovered individuals, and about 50 susceptible individuals that have escaped infection.

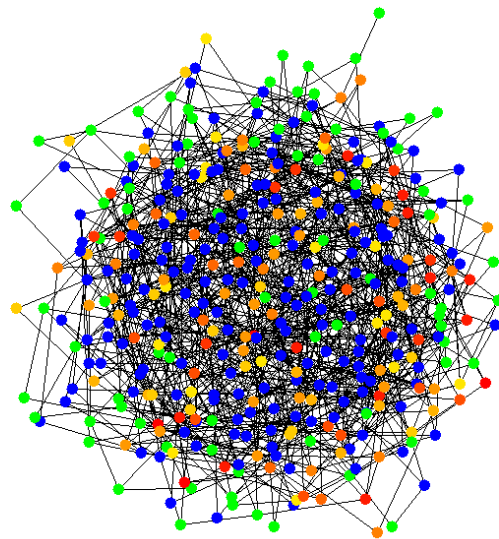


Figure 3: An example Random network during the simulation process

Figure 4 shows a different experimental run for the same network, with an infection peak occurring slightly later. This closely resembles Figure 1, where the infection peak occurs at time 34, although the peak is slightly lower (122 individuals), while the total number of people infected is higher (378).

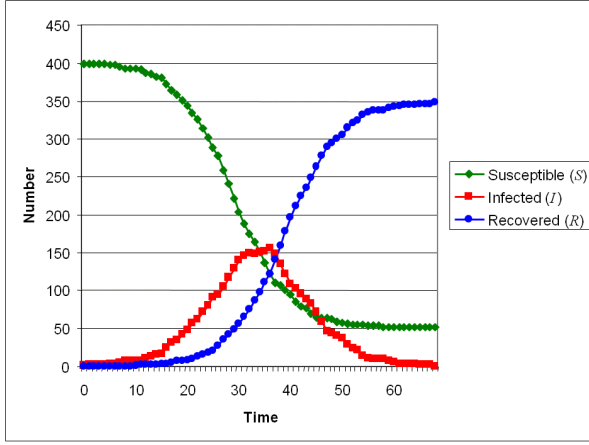


Figure 4: Agent-based simulation of disease spread for the Random network in Figure 3

For the Torus, in contrast, the number I of infected people remains roughly constant over time, and the total number infected increases approximately linearly, as the disease burns slowly around the Torus. Figure 5 illustrates this. The disease burns for longer, but infects fewer people than for the Random network, since it tends to die out early. For a network that was inherently linearly with time, so that the total number of individuals infected would increase quadratically – still much slower than the exponential growth with the Random network illustrated in Figure 4, and giving a smaller and later peak for I . We are particularly interested in the change in disease-spreading behaviour as the network transitions towards a Random topology.

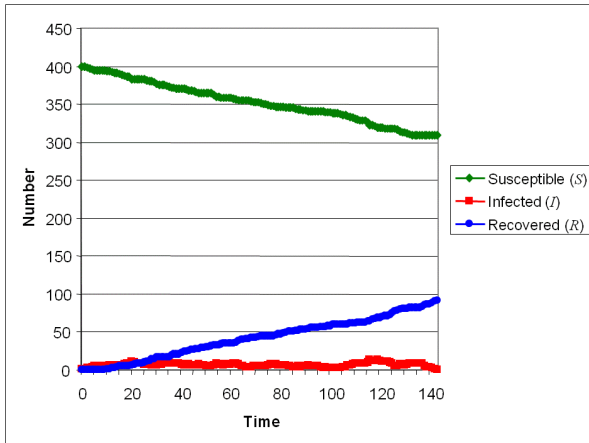


Figure 5: Agent-based simulation of disease spread for the Torus network

4. THE PANDEMIC THRESHOLD

When we examine the total number of people infected, shown in Figure 6, we can distinguish pandemics (where most of the population is infected) from disease outbreaks which die out early. Each data point in Figure 6 represents an average resulting from running the

disease simulation once each over 20 different generated networks for each value of the Kawachi parameter p .

For Small-World, Random, and Scale-Free networks, pandemics occur at about $\gamma > 0.4$. For the Torus network ($p = 0$), a higher value of about $\gamma > 0.6$ is required. This effect has also been reported by Watts and Strogatz [7].

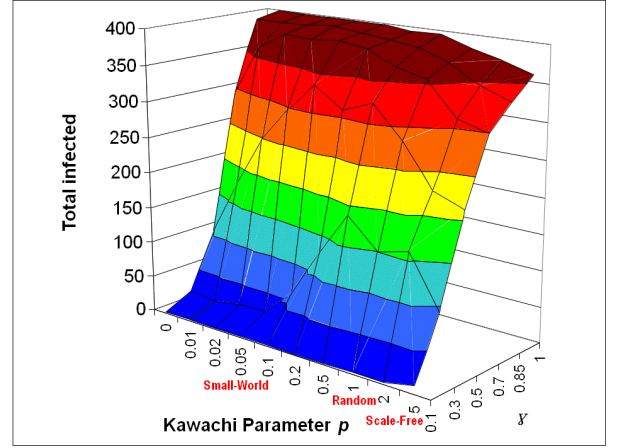


Figure 6: Total number of people infected for different networks and values of γ

For a network with average degree, or number of contacts per person, of d , we would expect the threshold to occur at $\gamma d = 1$. However, this ignores the fact that, in structured networks, many contacts are “wasted” (from the point of view of the disease organism) in that they lead to people that have already being infected by related individuals. Unstructured random networks transmit disease well since, from the point of view of the disease organism, no links are “wasted.” In Random networks of N people, the average distance D between people is approximately given by [8]:

$$D = \frac{\log N}{\log d} \quad (4)$$

We can invert this to define the effective degree d_{eff} of an arbitrary network by:

$$d_{\text{eff}} = N^{1/D} \quad (5)$$

This relates each network to an equivalent random network with similar disease-spreading ability. For the Random networks in our experiments, this effective degree is 5.5, while for the Torus network it is 1.3. If, as has been suggested [6], the human race has an average distance of 7 (corresponding to “six degrees of separation”), this would give the human network an effective degree of about 25.

Figure 7 illustrates the usefulness of this concept of effective degree – the transition to a pandemic occurs approximately when the effective reproduction number $\gamma d_{\text{eff}} = 1$ (corresponding to $\gamma = 0.2$ for the Random network, and $\gamma = 0.8$ for the Torus).

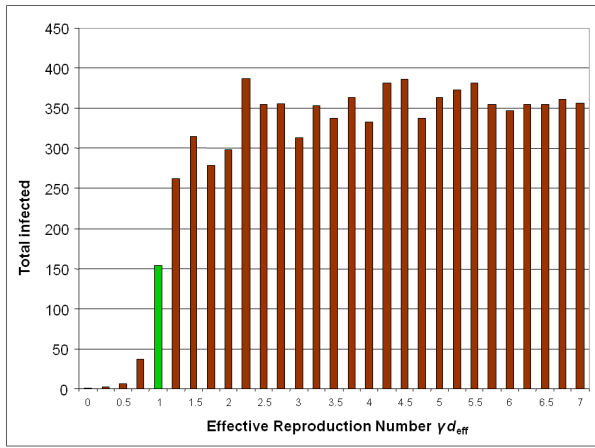


Figure 7: The pandemic threshold occurs at an effective reproduction number $\gamma d_{\text{eff}} = 1$

For comparison, using the actual degree, which is 4 for the networks we have used, gives a much less clear transition, as shown in Figure 8. This is because Figure 8 averages over different networks without taking into account their shape, while Figure 7 uses the effective degree d_{eff} to adjust for different network structures.

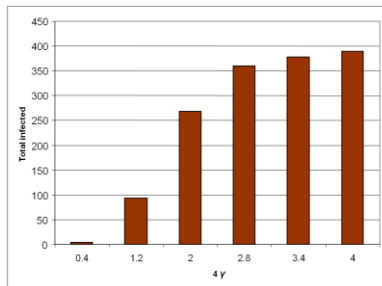


Figure 8: Replacing d_{eff} in Figure 7 by the actual degree of 4 gives a much less clear transition

A closer look at the right-hand side of Figure 6 shows that in the Scale-Free network for γ around 0.5, fewer people are infected on average than for the Random network (statistically significant at the 10^{-3} level). Scale-Free networks have most nodes with degrees below average, balanced by a few highly connected “hubs” having degrees very much more than average. However, the initially infected individual is likely not to be a “hub,” and the disease may therefore die out before a “hub” becomes infected, thus avoiding a pandemic. Consequently, at least for relatively small Scale-Free networks, the critical value of γ must be slightly higher than for the Random case.

However, the bottom right corner of Figure 6 shows that Scale-Free networks are characterised by an initial increase in the number of infected individuals (an epidemic in the mathematical sense) even for very low values of γ , in agreement with theoretical predictions [9].

At higher values of γ (about 0.7 and above), the networks in which disease spreads to the most people are Small-World networks with $p = 0.05$, and transitional networks with p ranging up to 0.5. These networks have a high average number of distinct paths between nodes [4]. Since previously infected and

recovered individuals “block” the flow of disease, the presence of multiple distinct paths allows the disease to spread more efficiently. This effect also means that the disease spreads more efficiently on the Torus network than would have been predicted simply on the basis of its effective degree.

5. THE INFECTION PEAK

Figure 9 shows the peak number of infected individuals I for varying networks and values of γ . The highest peaks occur for Random and Scale-Free networks with high values of γ . Empirically the peak value can be predicted by the formula $53 \gamma d_{\text{eff}} - 16$ (with a correlation of 0.86, $R^2 = 0.75$).

In comparison, for the corresponding SIR equations (with $\beta^* = 1/\delta$ and $\alpha^* = \beta^* \gamma d_{\text{eff}}/400$), the peak value is slightly lower, approximately $43 \gamma d_{\text{eff}} - 22$ (with a correlation of 0.98, $R^2 = 0.96$). In general, incorporating the effective degree d_{eff} into the SIR equations provides an approximation to the spread of disease on a network, but a relatively crude one. More realistic prediction of disease spreading requires agent-based simulation.

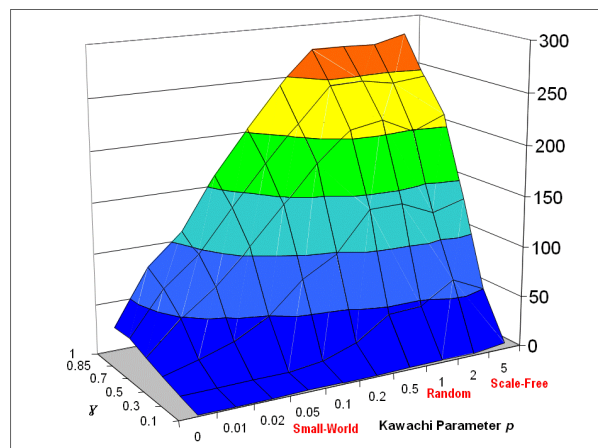


Figure 9: Number of infected individuals I at infection peak

6. THE TIME FACTOR

Figure 10 shows that the time to reach the peak infection is on average highest when the effective reproduction number $\rho = \gamma d_{\text{eff}}$ is just above 1. This is because when ρ is less than 1, the exponential growth of a pandemic does not occur, while if $\rho = \gamma d_{\text{eff}}$ is much greater than 1, the pandemic occurs very quickly.

We can define a piecewise linear function λ of ρ , based on Figure 10, which captures approximately the nonlinear effect of ρ on the time to reach the infection peak.

$$\lambda = \begin{cases} \rho, & \rho \leq 1.25 \\ 0.75 + \frac{3.25 - \rho}{4}, & 1.25 \leq \rho < 3.25 \\ 0.75, & \rho > 3.25 \end{cases} \quad (6)$$

For values of ρ greater than 1.25, λ substitutes values from the range 0.75 to 1.25 which have a similar effect on the time to reach the infection peak. Consequently, we can use λ , in conjunction with other variables, to predict the time to reach peak infection.

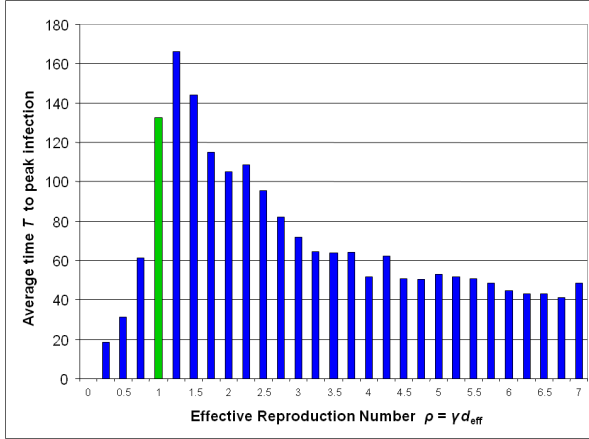


Figure 10: The greatest average time T to peak infection occurs when the effective reproduction number $\rho = \gamma d_{\text{eff}}$ is just above 1

The time to reach the peak also increases with the duration δ of the infective period, and the average distance D between people (Watts and Strogatz [7] also report this D effect). However, the time T to reach the infection peak can be predicted empirically without direct reference to D , using the formula:

$$T = 0.73 \delta^{0.80} 7.9^\lambda - 1 \quad (6)$$

which fits the data with a correlation of 0.90 ($R^2 = 0.81$), as illustrated in Figure 11.

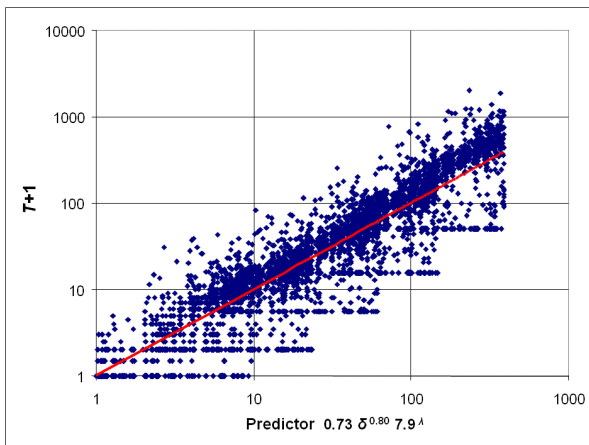


Figure 11: Empirical prediction of time T to peak infection by $0.73 \delta^{0.80} 7.9^\lambda - 1$

7. MORE REALISTIC SIMULATIONS

The simulations presented here permit an exploration of the important impact that network effects have on the spread of disease. However, more realistic agent-based

simulations [3] will also take into account the daily activities of people of different ages. People move between work, home, and school, and this movement is influenced by the presence or absence of illness. The number of social contacts that people have also varies in complex ways.

Developing more realistic simulations permits even better estimations of the infection peak and of the total number of people affected, but also requires realistic social network structures [8] in order to generate accurate results.

8. DISCUSSION

Agent-based models are useful tools for predicting the spread of new diseases within a population. The behaviour of such models is dependent on the topology of the social network connecting people. Using network rewiring and agent-based simulation techniques, we have investigated the network attributes which influence the spread of disease. In particular, we have identified the concept of “effective degree” as being of particular predictive value.

We have looked at three key disease characteristics: the peak number of infected individuals, the time to reach this peak infection, and the total number of individuals affected by the disease. All three significantly influenced by the effective degree of the network. Consequently, more realistic agent-based simulations [3] should pay particular attention to accurately modelling the effective degree of real-world social networks.

However, the modelling we have conducted has wider implications: the susceptible (S), infected (I), and recovered (R) states can be related to the spread of ideas and social trends through an organisation or society. The states S , I , and R can correspond to individuals who are *unaware* of a new idea, *enthusiastic* about it, and *disillusioned* by it. Ideas and social trends may therefore spread more effectively in an organisation with high effective degree (i.e. one with a low average network distance between people).

9. ACKNOWLEDGEMENTS

Discussions with Alex Skvortsov, and comments on an earlier draft of this paper by Peter Dawson, are gratefully acknowledged.

REFERENCES

1. Barry, J. (2005), *The Great Influenza*, Penguin: London.
2. Giesecke, J. (2002), *Modern Infectious Disease Epidemiology*, 2nd edition, Arnold: London.
3. Skvortsov, A.; Connell, R.; Dawson, P. & Gailis, R. (2007), “Epidemic Modelling: Validation of Agent-based Simulation by Using Simple Mathematical Models,” *MODSIM 2007 International Congress on Modelling and Simulation*.

4. Dekker, A. (2007), "Studying Organisational Topology with Simple Computational Models,"
5. Kawachi, Y.; Murata K.; Yoshii, S. & Kakazu Y. (2004), "The structural phase transition among fixed cardinal networks." *Proc. 7th Asia-Pacific Conf. on Complex Systems*, 247–255. Australia: Central Queensland University.
6. Watts, D. (2003), *Six Degrees: The Science of a Connected Age*. London: William Heinemann.
7. Watts, D. & Strogatz, S. (1998), "Collective dynamics of 'small-world' networks," *Nature*, Vol. 393, 4 June, 440–442.
8. Dekker, A. (2007), "Realistic Social Networks for Simulation using Network Rewiring," *MODSIM 2007 International Congress on Modelling and Simulation*.
9. Pastor-Satorras, R. & Vespignani, A. (2001), "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, Vol. 86, No. 14, 3200–3203.