Physica A 389 (2010) 561–576

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Stochastic epidemics and rumours on finite random networks

Valerie Isham [a], Simon Harden [a,*], Maziar Nekovee [b,c]

[a] Department of Statistical Science, University College London, London, UK

[b] Centre for Computational Science, University College London, Gordon Street, London WC1H 0AJ, UK

[c] Mobility Research Centre, BT, Polaris 134, Adastral Park, Martlesham, Suffolk IP5 3RE, UK

ARTICLE INFO

Article history:
Received 19 May 2009
Received in revised form 24 September 2009
Available online 7 October 2009

Keywords:
Epidemic models
Rumour models
Random networks
Stochastic models

ABSTRACT

In this paper, we investigate the stochastic spread of epidemics and rumours on networks. We focus on the general stochastic (SIR) epidemic model and a recently proposed rumour model on networks in Nekovee et al. (2007) [3], and on networks with different random structures, taking into account the structure of the underlying network at the level of the degree–degree correlation function. Using embedded Markov chain techniques and ignoring density correlations between neighbouring nodes, we derive a set of equations for the final size of the epidemic/rumour on a homogeneous network that can be solved numerically, and compare the resulting distribution with the solution of the corresponding mean-field deterministic model. The final size distribution is found to switch from unimodal to bimodal form (indicating the possibility of substantial spread of the epidemic/rumour) at a threshold value that is higher than that for the deterministic model. However, the difference between the two thresholds decreases with the network size, n, following a $n^{-1/3}$ behaviour. We then compare results (obtained by Monte Carlo simulation) for the full stochastic model on a homogeneous network, including density correlations at neighbouring nodes, with those for the approximating stochastic model and show that the latter reproduces the exact simulation results with great accuracy. Finally, further Monte Carlo simulations of the full stochastic model are used to explore the effects on the final size distribution of network size and structure (using homogeneous networks, simple random graphs and the Barabasi–Albert scale-free networks).

1. Introduction

In the 21st century, there are many infectious diseases that pose substantial threats to human and animal populations. There are concerns over the spread of HIV/AIDS, the emergence of new infections such as SARS, the threat of a new human influenza pandemic arising from avian influenza, possible outbreaks of smallpox resulting from terrorist action, and many others. Mathematical models have an important role to play in controlling the spread of such infections. In addition, the infectious agents and the infected hosts need not be biological systems. Individual personal computers and servers are regularly targeted by viruses spread across computer networks, while the effective transmission of information over the internet and the behaviour of social interaction networks are topical research interests. In many ways, the spread of information resembles that of infection, and the models that have been developed have many features in common.

In each case, individuals are classified as being of one of three types: susceptible, infected or removed for infections; ignorant, spreader or stifler for information or rumours. The simplest models assume homogeneous mixing of hosts, so that susceptibles become infected at a rate proportional to the current numbers of susceptibles and infectives, and similarly for

* Corresponding author. Tel.: +44 2076791872; fax: +44 2073834703.
E-mail address: simon@stats.ucl.ac.uk (S. Harden).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.10.001

the rate of spread of information from spreaders to ignorants. For the infection process, the only other transition is from infective to removed (e.g. through recovery, quarantine or death) which is generally assumed to take place at a constant rate per infective. For the information process, similar transitions (from spreader to stifler) at a constant rate per spreader may also be allowed, but in addition such transitions may occur if a spreader contacts another spreader or stifler (who is already in possession of the information or rumour and so discourages the spreader). Thus, the 'stifling' rate is nonlinear and this leads to interesting differences between some properties of epidemics and rumours.

Mathematical models have many uses. They may be used for careful exposition of issues and gaining theoretical understanding, for example, of the role played in the spread of infection or information by heterogeneity of hosts with regard to their behaviour or susceptibility. They can be used to answer specific questions of interest, such as the effect of a particular control strategy, or to reach qualitative conclusions that distinguish those factors and sources of heterogeneity largely responsible for driving the dynamics from those that have little influence. In turn, these may help in formulating what data are needed before parameters can be determined and the models used for prediction. Simple models give robust approximations, while more complex models require more detailed assumptions (and tend to have more parameters). In real time, models may sometimes be used simply as a means of summarising large quantities of information about the dynamics of an outbreak.

Understanding the mechanisms by which an infection spreads brings possibilities for its control. Questions include how to detect a new outbreak of an existing infection or the emergence of a new one, what action is needed to prevent its spread, and what contingency plans are needed for its practical implementation? For a recurrent infection, what/when/how should a routine strategy (*e.g.* vaccination) be implemented? On the other hand, rather than trying to limit the spread of infection, the aim may be to find the most efficient way to spread information or a rumour quickly over a computer network. For instance, large organisations might wish to distribute a critical anti-virus software upgrade over corporate networks or a viral marketing campaign may want to spread a video quickly over the Internet. Rumour-like mechanisms form the basis of an important class of data dissemination algorithms in distributed computer systems. These algorithms are generally known as *gossip* protocols [1]. The principle underlying these algorithms mimics the spread of rumour among humans or epidemics in populations. A process that wishes to disseminate a new piece of information to the system does not send it to a server, or a cluster of servers, in charge of forwarding, but rather to a (randomly chosen) set of other peer processes that it *knows*. In turn each of these processes does the same, forwarding the information to some of its peers.

The most fundamental question for the spread of infection is, if a single infectious individual is introduced into a closed and susceptible population, under what conditions (summarised as a *threshold theorem*) will the infection spread to infect a large proportion of the population, as opposed to dying out without having done so? If it does spread, then what is the distribution of the *final size* of the outbreak, that is, the total number infected? Similar questions are equally fundamental to the spread of information or a rumour. Analytic results are usually asymptotic, applying in the limit as the population size becomes arbitrarily large, and so the question arises as to how these thresholds behave for finite populations. In addition, threshold theorems are most easily obtained for a homogeneously mixing population in which contacts between every pair of individuals are equally likely, and there is considerable interest in extending these results to structured populations, either to metapopulations consisting of subgroups where there is homogeneous mixing within and between subgroups, or to populations with a network structure. Related issues arise in other fields and there is a long history of research on such questions in the statistical physics literature; see Ref. [2] for an extensive review.

With the growing interest in complex networks, the question of the behaviour of epidemic and rumour models on such structures has become the subject of much recent research. In particular, a number of recent studies have shown that introducing the complex topology of the social networks along which the spreading takes place can greatly impact on both the threshold behaviour and the dynamics of these models [3–6]. The majority of previous studies have their focus on the *infinite population* limit for which statistical fluctuations were expected to become negligible, hence justifying a deterministic treatment which focuses on obtaining the mean values of the quantities of interest.

However, even with large populations fluctuations do not always average out to result in a small overall effect. Furthermore, even if deterministic models give the mean behaviour of the corresponding stochastic system, in application to finite systems it is important to take the variability of individual realisations into account and, for example, to have information on the probability distribution of the number of individuals within a society who hear a certain rumour after its spreading, rather than only the mean number. Finally, the threshold concept that has been used so far in both epidemic [4,5,7] and rumour dynamics [3] is based on bifurcations caused by nonlinearities in the deterministic mean-field models. A suitable extension of the threshold concept to stochastic epidemics is currently missing.

The focus of the current paper is, therefore, to investigate in detail such stochastic aspects of epidemic and rumour models on networks, paying particular attention to the relationship between the threshold concept and the final size of the epidemic/rumour in the stochastic versus deterministic models.

In Section 2 we first briefly review the necessary background on stochastic epidemic and rumour models for homogeneously mixing populations, and discuss some simple models for random networks. Then, in Section 3, we consider a stochastic model for a rumour spreading on the nodes of a random network; the corresponding epidemic special case is obtained by setting one parameter to zero. An approximation is proposed in which the correlation structure of the network is retained but dependence between the infection status of neighbouring nodes is omitted. The resulting model is simple to analyse mathematically and its properties, including thresholds and final size distributions, can be explored numerically and related to those of a deterministic approximation discussed in an earlier paper, Ref. [3]. In particular, we introduce the

concept of a threshold in the context of stochastic epidemic and rumour models [8] on networks, and explore its dependence on the population size. We show that in the case of finite networks the thresholds for the stochastic models are higher that their deterministic counterparts, but the difference decreases with the network size, n, following a $n^{-1/3}$ asymptotic behaviour. Furthermore, we find that fluctuations in the final size of the epidemic are retained as the network size increases such that even in the infinite system size limit the deterministic model greatly overestimates the mean of the final size of the epidemic

Finally, in Section 4, Monte Carlo simulation results for the full stochastic model which incorporate the complete network structure are presented, as well as comparison with the approximate stochastic model when this dependence is ignored.

2. Background

2.1. The SIR epidemic model

Mathematical models of epidemics have a long and successful history, going back almost 250 years [9]. In the simplest models, hosts are divided into one of three categories: *susceptible, infected, removed*, leading to the term *SIR model*. Removed individuals play no part in the transmission of infection, they might be quarantined or recovered and immune. This partition is appropriate for microparasitic infections such as bacteria and viruses, where for most purposes the parasite can be assumed to replicate rapidly within the host to reach an equilibrium level (in contrast to macroparasitic infections such as helminths which require more complicated models because part of the parasite lifetime is external to the host, whose parasite load therefore only builds up through reinfection). In this paper, we consider a closed population of size n and denote the random numbers of susceptibles, infectives, and removals at time t by $X(t)$, $Y(t)$ and $Z(t)$ respectively, using $x(t)$, $y(t)$ and $z(t)$ to denote the corresponding realisations.

In Ref. [9], in a deterministic analysis where $x(t)$, $y(t)$ and $z(t)$ are treated as continuous variables, the assumption was made that the rate of new infections is simply a constant multiple, $\beta x(t)$ say, of the current number of susceptibles. After Daniel Bernoulli, it was another hundred years or so before the physical basis of the causes of infectious disease became established and mathematical modelling of transmission dynamics developed much further. The principle of *homogeneous mixing* (mass action) was established [10–12], which assumes that the infection rate is proportional to both the number of susceptibles *and* the number of infectives. Thus, in a deterministic model

$$\begin{aligned} \mathrm{d}x(t)/\mathrm{d}t &= -\beta x(t)y(t)/n \\ \mathrm{d}y(t)/\mathrm{d}t &= \{\beta x(t)/n - \delta\}y(t) \\ \mathrm{d}z(t)/\mathrm{d}t &= \delta y(t) \end{aligned} \tag{1}$$

where β can be interpreted as the rate at which infectives make potentially infective contacts (and $x(t)/n$ represents the chance that such a contact is with a susceptible) and δ is the rate at which infectives recover, corresponding implicitly to an infectious period that is exponentially distributed with mean $1/\delta$ in the stochastic version of the model. The scaling of the infection rate by the population size, n, guarantees that the number of potentially infective contacts per unit time remains fixed as the population size increases.

It can be seen from (1) that at the beginning of an outbreak the number of infectives increases as long as

$$x(0)/n > \delta/\beta = 1/R_0, \tag{2}$$

where the *reproduction ratio* R_0 can be interpreted as the mean number of susceptibles directly infected by a single infective during their infectious period if all contacts are with susceptibles. Thus there is a threshold at $R_0 = 1$, with the characteristic epidemic curve whereby the number of infectives $y(t)$ increases to a maximum and then decreases, applying only if $R_0 > 1$. Otherwise, the outbreak simply dies away. In an endemic situation, where there is an open population with recruitment of susceptibles (such as applies for many childhood infections), (2) underpins the vaccination strategy that aims to protect the whole population from major epidemic outbreaks by keeping the susceptible proportion of the population below $1/R_0$ (*herd immunity*). It can be shown [12] that, as $t \to \infty$, the number of removals satisfies the *final size equation*:

$$n - z(\infty) = x(0)\mathrm{e}^{-z(\infty)R_0/n}. \tag{3}$$

In the stochastic SIR model, the variables $X(t)$, $Y(t)$, $Z(t)$ and their corresponding values $x(y)$, $y(t)$, $z(t)$ are treated as integers. The model is a Markov model, so that it is specified by its transition probabilities. Thus, given $X(t) = x(t)$, $Y(t) = y(t)$, $Z(t) = z(t)$, transitions to $x(t) - 1, y(t) + 1, z(t)$ or to $x(t), y(t) - 1, z(t) + 1$ occur at *rates* $\beta x(t)y(t)/n$ and $\delta y(t)$ respectively (*i.e.* as $\mathrm{d}t \to 0$, the transitions occur in $(t, t + \mathrm{d}t)$ with probabilities $\beta x(t)y(t)\mathrm{d}t/n + o(\mathrm{d}t)$ and $\delta y(t)\mathrm{d}t + o(\mathrm{d}t)$ respectively). It is important to emphasise that the deterministic curve is *not* the mean of the stochastic epidemic. In particular,

$$\mathrm{dE}(Y(t))/\mathrm{d}t = \{\beta \mathrm{E}(X(t))/n - \delta\}\mathrm{E}(Y(t)) + \beta \mathrm{cov}(X(t), Y(t))/n \tag{4}$$

so that, in comparison with (1) it can be seen that the stochastic mean $\mathrm{E}(Y(t))$ would only be the same as the deterministic $y(t)$ if $\mathrm{cov}(X(t), Y(t)) = 0$. As the population size is fixed, this covariance is negative and, at the start of an epidemic, the stochastic mean will grow more slowly than its deterministic counterpart.

For the stochastic epidemic model, a branching process approximation for the initial stages of the outbreak (effectively an asymptotic approximation—in n—in which it is assumed that all contacts of infectives are with susceptibles) gives a threshold at $R_0 = 1$ for the behaviour of the outbreak [13]. For $R_0 > 1$, there is a major outbreak with probability $1 - R_0^{-Y(0)}$. Otherwise, or if $R_0 \leq 1$, there is only a minor spread of infection.

Informally, in a minor outbreak, only an asymptotically negligible proportion of the population is infected so that, for $R_0 \leq 1$, the final size distribution is J-shaped and has a single mode at 0. In contrast, for $R_0 > 1$, the final size distribution is U-shaped, with modes at 0 and at the location of the major outbreak. For those realisations where a major outbreak occurs, a Central Limit effect applies, and asymptotically the number of infectives is a Gaussian diffusion about the deterministic solution [14–21] in which the mean and variance scale with n. The duration of a major outbreak has been shown [22] to be $O(\ln n)$.

Thus the importance of allowing appropriately for stochasticity in epidemic models is clear: even in large populations there is uncertainty and the variation of individual realisations of an epidemic about the deterministic solution is of order $\sqrt{n}$. But the course of an outbreak is often determined by stochastic effects near the boundary of the state space where the number of infectives is small and the discreteness of the population is crucial. In the stochastic case, the infection is certain to die out eventually, after which further outbreaks can only occur if the infection is reintroduced into the population. In contrast, in the deterministic model, the number of infectives, $y(t)$, is always positive (assuming $y(0) > 0$), so that a second wave of infection can occur if there is a build up of susceptibles in an open population [23].

2.2. Rumour models on homogeneously mixing populations

Rumour models add a further feature to those present in epidemic models, namely stifling: when an attempt is made to spread a rumour to someone who has already heard it, then the spreader or the spreader and the listener may become removed (or stiflers, in the rumour terminology). Thus, rumour models can be viewed as a generalisation of epidemic models. A number of specific rumour models have been proposed and explored but the literature is considerably less extensive than that for epidemics. An introduction is given in Ref. [24] which reviews earlier work (including Refs. [15,25,26]) on a number of rumour models, and discusses their properties.

Here, we define a simple Markov rumour model, extending the discussion of the previous section. In a rumour context, the population is partitioned into *ignorants, spreaders* and *stiflers*, corresponding to the susceptibles, infectives and removals of the epidemic case; we retain the notation X, Y, Z for the numbers in the three categories, and assume that the population is closed and of size n. The 'infection' rate is as before, with ignorants becoming spreaders at a *per capita* rate $\beta Y(t)/n$. However, in rumour models, a spreader may be stifled (*i.e.* decide to stop spreading the rumour) if they contact another spreader or a stifler (who has already heard the rumour). We assume a homogeneous mixing form for the contact, and assume that in such a contact the spreader is stifled with a probability p, independently between contacts. We also allow spreaders spontaneously (without any form of contact) to become stiflers at a constant rate, δ; this is sometimes termed 'forgetting'. Thus we will assume that the total per capita rate at which a spreader becomes a stifler is $p\beta\{Y(t)+Z(t)\}/n+\delta$. Note that this model subsumes the SIR epidemic model, which is obtained by setting $p = 0$.

As with the epidemic models, the properties of interest focus on the conditions under which the rumour will spread and, if it does, the number of those that will ultimately hear it. The directed form of stifling contacts defined above is due to Ref. [27]. An undirected alternative is that in a contact between two spreaders, either one or both of the spreaders may be stifled. This approach was introduced in Refs. [15,24] focus their discussion on such a model with undirected stifling in which, in a contact between two spreaders, both are stifled, and where there is no forgetting ($\delta = 0$). However, model variants are discussed briefly and, in particular, they note that forgetfulness is a necessary requirement for the existence of threshold phenomena. Intuitively, stifling only occurs as a result of contact between a spreader and another spreader or stifler, and is in some sense a second-order phenomenon, while forgetting is a first-order effect. At the start of a rumour, almost all contacts are between spreaders and ignorants, so that stifling cannot by itself act as a break on the spread of the rumour, while forgetfulness can dominate the spreading if δ is sufficiently large.

The rumour and epidemic models described so far assume that the principle of homogeneous mixing applies. Much of the recent work on rumours has not assumed this and deterministic analyses, for instance, Refs. [3,1,6,28,29], often focus on specific examples based around social and computer networking.

2.3. Embedded Markov chains

Continuous time Markov chains, such as the SIR epidemic model or one of the rumour models described above, are usually defined in terms of rates of transition (q_{ij}) between distinct states $i, j \in S$, where S is the state space of the system. If the probabilities $(p_{ij}(t))$ of transition between distinct states at time t are differentiable at $t = 0$, then $p_{ij}(t) = q_{ij}\,t + o(t)$ as $t \to 0$, and the properties of the system are determined by the transition rates. In particular, the successive durations of stay between transitions are independent and exponentially distributed random variables. If the system is in state i, then the duration of stay is exponentially distributed with parameter $\sum_{j\in S, j\neq i} q_{ij}$ and, conditionally upon a transition taking place, the transition is to state j, $(j \neq i)$ with probability $q_{ij}/\sum_{j\neq i} q_{ij}$. The sequence of states is itself a Markov chain in discrete time, known as the *embedded Markov chain*, with transition matrix $(q_{ij})_{ij\in S}$. The diagonal elements, $\{q_{ii} : i \in S\}$, of this

transition matrix are all zero, since it is not possible for two successive durations of stay of the continuous time process to be in the same state. The embedded Markov chain provides a convenient route to the simulation of the continuous time process, as it is only necessary to simulate the sequence of independent, exponential durations of stay together with the Markov sequence of states. This method of simulation is sometimes known as the Gillespie algorithm [30]. Properties of a continuous time Markov chain, such as the final size distribution of an epidemic, for which the temporal durations of stay in the sequence of states are irrelevant, are often most easily found in terms of the embedded Markov chain, and we shall use this approach in Section 3.2 to determine the final size distribution for some epidemic and rumour models.

2.4. Random networks

In the preceding sections, epidemic and rumour models have been defined for a homogeneously mixing population in which contacts between all pairs of hosts are equally likely. However, for many applications it is more appropriate to consider a structured population in which the rates of contact may vary between pairs, and then a metapopulation model may be used (see for example Ref. [31] where there are two levels of mixing, with local and global contacts). Alternatively, and in the context to be considered below, contacts may only be possible between each host and a group of neighbours, so that the underlying population has a network structure. Models for some simple random networks (see Ref. [32], and also Ref. [33] for an application-led review) that will be used in this paper are described briefly below. In each case we assume that the network has n nodes (hosts) and give a rule for joining neighbouring nodes by edges to form a graph.

Simple random graph. In this model, each pair of nodes is connected by an edge with probability π, independently between pairs. Thus the *degree* of an arbitrary node (*i.e.* the number of nodes to which it is connected) has a binomial distribution with index $n-1$ and mean $(n-1)\pi$, and for large n is approximately Poisson distributed. This is an example of an *uncorrelated* graph, where the degrees of neighbouring nodes have zero correlation over their remaining edges.

An uncorrelated graph with an arbitrary degree distribution. In this case, random degrees from some chosen distribution are allocated independently to the nodes, and it is assumed that the sum of these degrees is even (if not, a degree can be resampled to ensure this condition). A node with degree k can then be regarded as having k arms, so that an uncorrelated graph can be constructed by pairing the arms between the nodes at random. Such a graph is sometimes known as a Molloy–Reed graph [34]. A *homogeneous* random graph is a special case in which all nodes have the same fixed degree.

Scale-free graphs. A scale-free graph is one in which the degree distribution has a power-law tail. One example is the Yule–Simon distribution [35], which has a probability mass function $f_D(k) = (\rho-1)B(k,\rho)$, where $B(\cdot,\cdot)$ is the Beta function. For large k, $f_D(k) \sim k^{-\rho}$; for $\rho > 2$, the mean of the distribution is $(\rho-1)/(\rho-2)$. The degree distributions of many empirical networks can be well approximated by distributions with power-law tails, and values of the negative exponent, ρ, between 2 and 4 are typical [33]; such degree distributions will have infinite variance. An uncorrelated graph with this degree distribution can be constructed via the 'pairing arms' algorithm. Scale-free networks are characterised by the presence of *hubs*, *i.e.* nodes with very high degrees, a feature typical of computer networks.

The Yule–Simon distribution was originally put forward as the equilibrium distribution of a particular stochastic process called a *preferential attachment process*. The example given in Ref. [35] is of a model for writing a book, where words are added one a time. At each step, with a constant probability the word is a new word that has not so far appeared, while with the remaining probability the word is sampled from the existing words in the document in proportion to their current frequencies. A rather similar construction of a random network was proposed in Ref. [36] as a model for network growth. Initially a small number of nodes are completely connected. Nodes are then added one at a time, each with the same fixed number (ω say) of edges that are used to connect it to the existing nodes in proportion to their current degrees. The Barabasi–Albert construction leads to a scale-free network where the degree distribution is such that for large k, $f_D(k) \sim k^{-3}$.

3. Rumours and epidemics on networks

3.1. Model specification and an approximation

In the rest of this paper, we will consider the rumour model of Section 2.2 running on a population for which, rather than being homogeneously mixing, the contact structure is given by a random network as described above. One application is to online social interaction networks [37], where typically there might be $O(10^6)$ nodes and a very skew degree distribution. The network structure is fixed and does not evolve in time. The differences from the case with a homogeneously mixing population are that a spreader node can only pass on the rumour or be stifled by a neighbouring node, as specified by the network, and that spreaders now contact *each* neighbour at a rate λ so that the total rate at which nodes contact their neighbours varies from node to node. For internet rumours, this assumption is an approximation to the practical situation where rumours are broadcast simultaneously to all neighbouring nodes. If all nodes have the same degree (k) then the correspondence with the homogeneously mixing population is that $\beta = \lambda k$.

We assume that there are n_k nodes of degree k and that of these X_k, Y_k and Z_k are respectively ignorants, spreaders and stiflers (where the dependence on time is omitted). Since $Z_k = n_k - X_k - Y_k$, the values of Z_k need not be specified in the following analysis. We denote the network degree–degree correlation function by

$$p_{jk} = P(\text{neighbour node has degree } k \mid \text{ index node has degree } j).$$

For example, in an uncorrelated graph, $p_{jk} \propto kp_k$, where $\{p_k\}$ is the marginal degree distribution. In Ref. [3] an analysis of this model is discussed in which the influence of the network structure is wholly encapsulated by the p_{jk} matrix. Specifically, $\lambda k X_k$ is the total rate at which degree k ignorants are contacted by their neighbours, and thus the total rate of transitions from (X_k, Y_k) to $(X_k - 1, Y_k + 1)$ can be written as

$$\lambda k X_k \sum_j p_{kj} A_{jk}$$

where A_{jk} is the probability that a node is a spreader given that it has degree j and is the neighbour of an ignorant node of degree k. In the [3] analysis, the probability A_{jk} is approximated by the proportion of degree j nodes that are spreaders so that the total transition rate is simply

$$\lambda k X_k \sum_j p_{kj} Y_j / n_j. \tag{5}$$

Thus the dependence of the rumour status of the degree j node on that of its neighbour is ignored. A similar approximation is made with regard to stifling so that the total rate of transitions from (X_k, Y_k) to $(X_k, Y_k - 1)$ is

$$Y_k \left\{ \delta + \lambda p k \sum_j p_{kj} [n_j - X_j]/n_j \right\}. \tag{6}$$

The Markov model with these transition rates will be termed the *approximate* stochastic model to distinguish it from the full network model. The approximate model retains the dependence on the network structure but not the dependence that this induces in the rumour status of neighbouring nodes. In that sense, this model has a homogeneous mixing population structure and, when $p = 0$, the approximate model reduces to the SIR epidemic model. Note that the approximate model allows 'self-stifling': for example, if, initially $X_k = n_k - 1$, $Y_k = 1$ and $X_j = n_j, j \neq k$, then the stifling rate is $\delta + \lambda p k p_{kk}/n_k$, whereas in the full model this rate is simply δ. The effect of this difference is minimal except for very small networks.

In Ref. [3] a deterministic analysis of the approximate model is discussed; it is also assumed that rumours are broadcast simultaneously from a node to all its neighbours, but this makes no difference to the deterministic results. In particular, for a homogeneous network where all nodes have degree k, they show that the proportion of the population eventually hearing the rumour satisfies a final size equation $z = 1 - \mathrm{e}^{-Rz}$ where

$$R = \lambda k(1 + p)/(\delta + \lambda k p) = (1 + p)/(\psi + p)$$

and $\psi = \delta/(\lambda k)$. This is consistent with the SIR final size equation given in Section 2.1 when $p = 0$. For the rumour to spread (non-zero z), we need $R > 1$ or, equivalently, $\psi < 1$. It can be seen that if $\delta = 0$, then $R > 1$ (unless $p = 0$) and there is no threshold for the spread of the rumour, confirming the result noted in Section 2.2 for the case with a homogeneously mixing population. If $\delta > 0$, then the rumour spreads as long as $\delta < \lambda k$ ($\psi < 1$) regardless of the value of p, showing that the threshold is the same as for the SIR epidemic model. This result is extended to an uncorrelated network and it is shown that, to leading order in p, for the rumour to spread the condition is that $\delta < \lambda(\mu_D + \sigma_D^2/\mu_D)$. Thus, the threshold is determined not just by the mean degree, μ_D, but by the variance σ_D^2, through the presence of the index of dispersion term σ_D^2/μ_D. For many scale-free networks, the variance will be infinite, in which case this condition is always met and the rumour will spread regardless of the value of δ.

3.2. Stochastic analysis of the approximate model for a homogeneous network

In this section, we determine the properties of the approximate model, with transition rates given by (5) and (6) in the special case when the underlying network of n nodes is homogeneous, so that all nodes have fixed degree k. In this case, we have a bivariate Markov process $\{X(t), Y(t)\}$ with spreading rate $\lambda k XY/n$ and stifling rate $Y\{\delta + \lambda p k(1 - X/n)\}$. To obtain the final size distribution it is easiest to use embedded Markov chain techniques (see *e.g.* Ref. [38] for the case $p = 0$). Thus, for $x \geq 0$, $y \geq 0$ and $x + y \leq n$, if the current state is (x, y), the time until the next transition occurs is exponentially distributed with rate parameter $u_{xy} = (\delta + \lambda p k)y + \lambda(1 - p)kxy/n$. Given that a transition occurs, it is to $(x - 1, y + 1)$ with probability

$$\phi_x = \frac{x}{(\psi + p)n + (1 - p)x}.$$

Alternatively, with probability $\overline{\phi}_x = 1 - \phi_x$, the transition is to $(x, y - 1)$. Notice that these probabilities are independent of y and depend only on the current number of ignorants.

Also, the deterministic condition that $\psi < 1$ for epidemic spread corresponds to $\phi_x > 0.5$ in the embedded Markov chain. In the embedded chain, the initial states are $x(0) = n - 1$, $y(0) = 1$, and the states $(x, 0)$ are absorbing. It is certain that the chain will eventually reach the state $(s, 0)$, for some $s \leq x(0)$. Absorption in state $(s, 0)$ gives a final size of $n - 1 - s$. The possible transitions can be seen in Fig. 1.

Let π_{xy} denote the probability that the embedded chain ever reaches state (x, y) so that $\{\pi_{n-1-s,0}, s = 0, 1, \ldots, n - 1\}$ is the distribution of the final size of the epidemic. It can be seen from the figure, and by conditioning on the last transition,

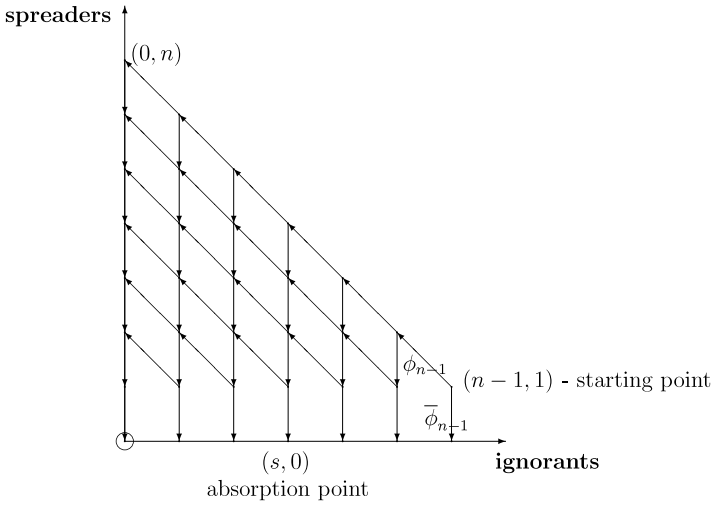


Fig. 1. Lattice of possible routes through ignorant/spreader state space.

that the probabilities π_{xy} satisfy the forward equations

$$\begin{aligned}
&\pi_{xy} = \pi_{x+1y-1}\phi_{x+1} + \pi_{xy+1}(1-\phi_x) \quad \text{for } x \geq 0, y \geq 2, x+y \leq n-1.\\
&\pi_{xy} = \pi_{x+1y-1}\phi_{x+1} \quad \text{for } 0 \leq x \leq n-2, y = n-x\\
&\pi_{x1} = \pi_{x2}(1-\phi_x) \quad \text{for } 0 \leq x \leq n-2,\\
&\pi_{x0} = \pi_{x1}(1-\phi_x) \quad \text{for } 0 \leq x \leq n-1,
\end{aligned} \tag{7}$$

with $\pi_{n-11} = 1$ given the initial condition.

The equations specified by (7) can be solved iteratively to get the final size distribution, starting from the initial state. Alternatively, a direct argument is as follows. The initial state is $(n-1, 1)$. Suppose that absorption takes place in state $(n-j, 0)$, so that the final size is $(n-1) - (n-j) = j-1$, and suppose for the moment we assume that $j \geq 3$. Then there have to be $j-1$ spreading steps, and j stifling steps, the last two of which must be from $(n-j, 2)$ and $(n-j, 1)$. The spreading probabilities depend only on the current value of x, so that the spreadings contribute a factor of $\prod_{i=1}^{j-1}\phi_{n-i}$ to the overall probability $\pi_{n-j\,0}$, while the final two stiflings each contribute a probability $\overline{\phi}_{n-j}$. It remains to determine the probability contributed by the other $j-2$ stiflings. Suppose these take place when $x = n-i_1, \ldots, n-i_{j-2}$, where $2 \leq i_1 \leq i_2 \ldots \leq i_{j-2} \leq j$. The $i_1, \ldots, i_{j-2}$ must satisfy additional constraints because the boundary state at $y = 0$ must not be entered and thus the number (y) of spreaders must be at least 2 when the stifling occurs. Suppose $x = n - i_\ell$. Then there have been $i_\ell - 1$ spreading events (resulting in increases in y) and $\ell - 1$ stiflings (decreases). Thus, given that initially $y = 1$, the stifling event can only occur if $i_\ell - (\ell - 1) \geq 2$, *i.e.* $i_\ell \geq \ell + 1$. Putting all this together, we obtain

$$\pi_{n-j0} = \left(\prod_{i=1}^{j-1}\phi_{n-i}\right)\overline{\phi}_{n-j}^2 \sum_{i_1=2}^{j}\sum_{i_2=\max(i_1,3)}^{j}\cdots\sum_{i_{j-2}=\max(i_{j-3},j-1)}^{j}\overline{\phi}_{n-i_1}\cdots\overline{\phi}_{n-i_{j-2}} \tag{8}$$

where this equation holds for $j \geq 3$. It is easily seen that for $j = 1, 2$, the corresponding probabilities are given by $\pi_{n-1,0} = \overline{\phi}_{n-1}$, and that $\pi_{n-2,0} = \phi_{n-1}\overline{\phi}_{n-2}^2$.

Note that, for computational purposes with large networks, it is considerably more efficient to obtain the final size distribution via the forward Eq. (7) than to use (8); the latter involves $O(4^n)$ operations, while the former requires only $O(n^2)$.

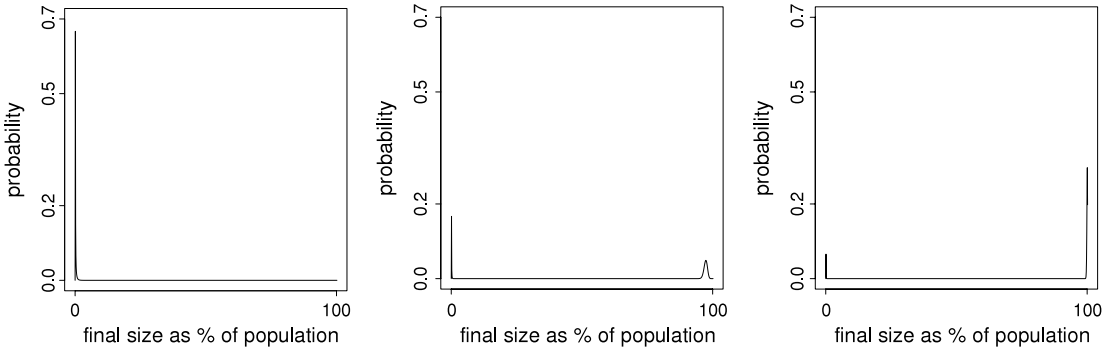
The time to extinction of the rumour can also be obtained, using a similar approach. Let T_{xy} be the time from the state (x, y) until the absorbing barrier is reached. Then

$$T_{xy} = \begin{cases} T_{x-1y+1} + \tau_{xy} & \text{with probability } \phi_x\\ T_{xy-1} + \tau_{xy} & \text{with probability } 1-\phi_x \end{cases}$$

where τ_{xy} is an exponentially distributed random variable with parameter u_{xy}, and all the distinct τ_{xy} are mutually independent. Thus, for example, the mean first passage times to the barrier satisfy the equations

$$\mathrm{E}(T_{xy}) = 1/u_{xy} + \phi_x\mathrm{E}(T_{x-1y+1}) + (1-\phi_x)\mathrm{E}(T_{xy-1})$$

for $x \geq 0, y \geq 0, x+y \leq n$, where we define $\mathrm{E}(T_{xy}) = 0$. This set of equations can be solved iteratively to find $\mathrm{E}(T_{n-11})$, the mean first passage time from the initial state $(n-1, 1)$. Writing $\mathrm{E}(T_{xy}) = E_{xy}$, it is easy to see that $E_{0y} = \sum_{j=1}^{y}(1/u_{0j})$.



(a) $R_0 = 0.5$, J-shaped. Final size will be small.

(b) $R_0 = 5$, U-shaped. Final size may be larger.

(c) $R_0 = 14$, U-shaped. Final size is likely to be large.

Fig. 2. Final size distribution and description for $p = 0.1$ and a range of values of R_0.

Then we may obtain successively $E_{11}, E_{12}, \ldots, E_{1,n-1}, E_{21}, E_{22}, \ldots, E_{2,n-2}$ and so on. Unfortunately, the results do not seem expressible in a simple algebraic form.

Other properties of the time to extinction, such as the variance or the moment generating function, can be found similarly.

3.3. Numerical results for the approximate model on a homogeneous network

We focus on the critical threshold for the spread of a rumour, as a function of the population size. For finite n and fixed p, we follow a suggestion in Ref. [8] for the SIR model and here define the critical threshold for the stochastic rumour model as the value of $R_0 = (1 + p)/(\psi + p)$ for which the final size distribution changes from J-shaped to U-shaped.

All the numerical results in this section are derived from a computer analysis built around the model described in Section 3.2. The forward Eq. (7) are used in preference to the analytical solution (8) as the number of calculations is $O(n^2)$ as opposed to $O(4^n)$. The results do thus not contain an element of error, and are approximations only to the extent that there is rounding to a finite number of decimal places. For instance, a network with 10,000 nodes requires the use of triple precision real numbers to calculate thresholds in order to prevent the appearance of 'ghost' peaks due to rounding errors.

Fig. 2 shows the final size distributions for a rumour model as a function of $R_0 = (1 + p)/(\psi + p)$. For $R_0 = 0.5$ the infection dies out quickly, for $R_0 = 5$ with high probability the infection dies out but with a small probability it spreads to a substantial proportion of the population, while for $R_0 = 14$ it is highly likely that almost all the population will be affected ($p = 0.1$ has been used in all cases). Very accurate numerical work is needed to detect the value of ψ for which the second peak exists, with the probability of a final size of m say, just exceeding those of $m + 1$ and $m - 1$.On the basis of numerical work for the SIR model with small populations, in Ref. [8] it was conjectured that the threshold value of R_0 has the form $R_0 \sim 1 + \kappa n^{-1/3}$, while Ref. [38] provided theoretical support via an approximation to the final size distribution as a mixture of the distribution for a small epidemic (a branching process) and that for a large epidemic (normal distribution).

Fig. 3 shows the threshold value of ψ plotted against p for a range of values of n up to $n = 10^4$. The values of ψ when $p = 0$ correspond to the threshold values of $R_0 = 1/\psi$ for the SIR epidemic model. For $p > 0$, the threshold values of R_0 are obtained by substituting for ψ in $R_0 = (1+p)/(\psi+p)$. It can be seen that, as $n \to \infty$, the threshold value of ψ approaches 1, with smaller values (and hence larger threshold values for R_0) as n decreases. As is to be expected intuitively, as p increases so that spreaders are increasingly likely to be stifled, the threshold values of ψ decrease (and corresponding R_0 values increase), but this effect dies out and the threshold becomes independent of p as population sizes increase. This matches the result for deterministic models described in Ref. [3]. We have not found an obvious explanation for the piecewise linear appearance of the threshold curve for $n = 10$. This effect is visible for small values of n, with the position of the elbow increasing with p as n increases (it is below $p = 0.1$ for $n = 5$, and around $p = 0.7$ for $n = 20$), but is no longer apparent well before $n = 50$.

It was conjectured in Ref. [8] that R_0 has the asymptotic form $1 + \kappa n^{-1/3}$. For $p = 0$, a nonlinear least squares (Gauss–Newton algorithm as implemented in R) fit of the form $R_0 = 1 + \kappa n^{\eta}$ to the values from $n = 10^3$ up to $n = 2 \times 10^5$ gives the fitted curve $R_0 = 1+0.879n^{-0.327}$, which is consistent with the conjecture, although it is clear that the convergence to the exponent of $-1/3$ must be very slow. The fitted slope was found to be very sensitive to the values of n included: excluding smaller values of n produced an exponent closer to $1/3$.

As an alternative approach, Fig. 4 shows a plot of $R_0 - 1 = (1 - \psi)/(\psi + p)$ against n on a log–log scale. Fitting straight lines, using a linear model of the form $\log(R_0 - 1) = A + B \log(n)$ to the data for $n = 10^3$ to $n = 2 \times 10^5$ leads to estimates of the coefficients given in Table 1. These results suggest some dependence of the intercept A on p for values of n in this range. The values of B show a small decrease with increasing p.

Fig. 5 shows the mean time until the rumour ceases to spread for $p = 0.5$ and two values of ψ. When $\psi = 0.5$ ($R_0 = 1.5$), the final size distribution is U-shaped, with positive probability of a substantial spread, while when $\psi = 1.0$ ($R_0 = 1$) and $\delta = 1.0$ the final size distribution is J-shaped. In both cases, the mean duration increases with $\log n$, which is consistent with the results of Ref. [22] for the SIR model, and would be expected to be similar for the rumour model.

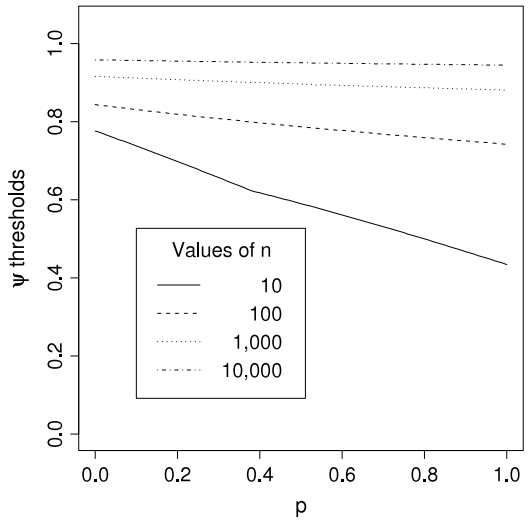


Fig. 3. Threshold points between unimodal and bimodal for different size networks.

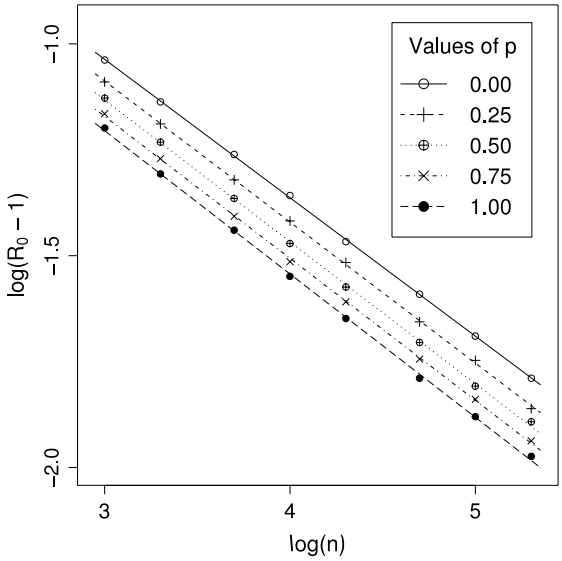


Fig. 4. Logs of population size and threshold: the points are exact and the lines are linear regressions on points for $\log(n) \geq 3$.

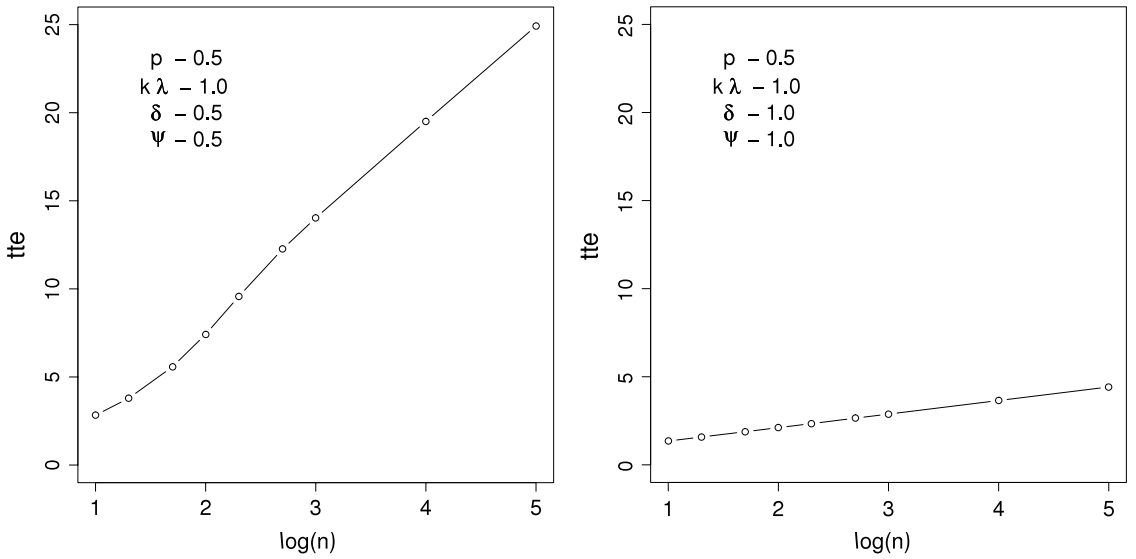
Table 1
Estimates of coefficients A and B in the log-linear form $\log(R_0 - 1) = A + B\log(n)$ for a range of values of p.

p	A	B
0.00	−0.056	−0.327
0.25	−0.089	−0.333
0.50	−0.127	−0.335
0.75	−0.167	−0.335
1.00	−0.190	−0.338

Fig. 6 shows the mean duration conditionally upon the maximum proportion of the population to whom the rumour is spread, for the case $\psi = 0.5$ ($R_0 = 1.5$). This figure shows clearly how, for larger networks, the mean duration increases very sharply once the maximum proportion exceeds the second peak of the final size distribution (here at about 0.6). For small n the process runs out of ignorants before large durations can build up.

3.4. Numerical comparison between stochastic and deterministic models

In this section, we compare numerically the properties of our approximate stochastic rumour model running on a homogeneous network, with those obtained by a deterministic analysis of this model [3] as described in Section 3.1.



(a) $R_0 = 1.5$, 2 peaks. (b) $R_0 = 1.0$, 1 peak.

Fig. 5. Mean time to extinction (tte) versus $\log(n)$ for rumours with different parameters.

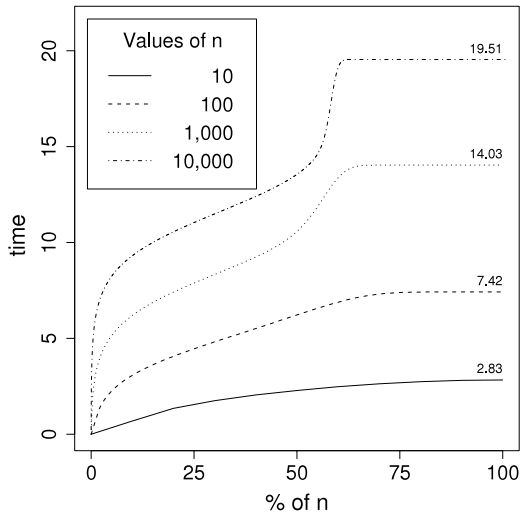


Fig. 6. Conditional times to extinction for $p = 0.5$, $\lambda k = 1$ and $\delta = 0.5$ so $R_0 = 1.5$ ($\psi = 0.5$).

For a network with 10,000 nodes, $p = 0.1$ and $\lambda k/\delta$ ranging from 0 to 50, the final size for the deterministic model is calculated and compared in Fig. 7 to two final sizes from the equivalent stochastic model, the overall average and the value at the second peak. The deterministic and stochastic second peak lines match almost exactly while the stochastic average final size is lower but approaches the other two lines asymptotically as $\lambda k/\delta \to \infty$. As discussed in Section 3.1, the deterministic model will only take off for $R > 1$ (equivalent here to $\lambda k/\delta = 1/\psi > 1$). Fig. 3 shows that the threshold point between unimodal and bimodal final size distributions in the stochastic case has an equivalent value of $1/\psi$ of about 1.03 for this network size and value of p. Above those values, the deterministic model is equivalent to a stochastic one conditional upon the rumour taking off and thus the coincidence of the deterministic final size and the stochastic second peak is expected. The stochastic average line has smaller final size for any value of $\lambda k/\delta$ as it will take into account the possibility that the rumour will not take off. As $\lambda k/\delta$ grows, that possibility becomes increasingly remote and all three lines will merge asymptotically. This comparison has been run for a range of network sizes and values of p, all giving a similar picture.

4. The full network model: Simulations and comparisons

In Sections 3.2 and 3.3 we discussed the properties of a stochastic model that can be regarded as an approximation to the full model formulated in Section 3.1 for a rumour evolving on a homogeneous random network. It was obtained by ignoring

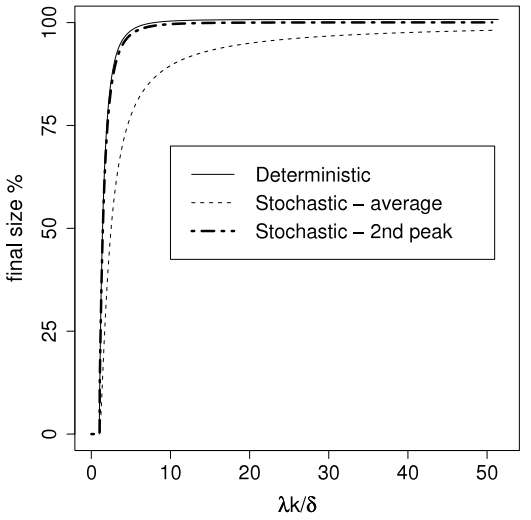


Fig. 7. Comparison of final sizes for deterministic and approximate stochastic models, $n = 10{,}000$, $p = 0.1$.

the dependence between the rumour states of neighbouring nodes, given their degrees. Intuitively, this approximation will have limited effect in the early stages when almost all nodes are ignorant of the rumour and will therefore have rather little effect on threshold results.

In this Section, we first explore the adequacy of the approximation by using simulations of the full model. As we are concerned only with thresholds and the final size distribution, and not with the time path of the dynamics, we are able to use the embedded Markov chain, as described in Section 3.2. We will then investigate the effects of different network sizes and structures on the thresholds for rumour spread and on the shape of the final size distribution. The network structures to be compared will be the homogeneous (fixed k) graph, the simple random graph and a scale-free graph obtained using the Barabasi–Albert algorithm (see Section 2.4). We will also look briefly at the effect of broadcast spreading versus independent spreading of the rumour to each neighbour separately. The problem in using simulation results is that to determine the threshold where the final size distribution changes from J-shape to U-shape, and hence when the probabilities cease to decrease monotonically, the final size distribution needs to be estimated extremely accurately. This requires very large numbers of realisations of the rumour spread to be simulated for each parameter set.

Only a subset of the possible comparisons will be reported here. As well as the network size (n) and structure, each network is governed by a single additional parameter (the degree k for the homogeneous graph, the edge probability π for the simple random graph, the number of edges ω added simultaneously in the Barabasi–Albert algorithm for the scale-free graph). The rumour evolution then requires three further parameters (the spreader contact rate λ, the stifling probability p, the forgetfulness parameter δ). The approximate model is determined by the composite parameter $\psi = \delta/(\lambda k)$ together with the stifling probability p. A thousand realisations have been used here for each choice of network and set of parameters. In each case, for a specific size and structure of network, only a single realisation of the random network has been used and the network is regarded as fixed (although results presented here have been confirmed by varying the initial network). The results would be affected by a further source of random variation if network variability is incorporated. For the 1000 simulations of the rumour process, the initial spreader is chosen independently and at random for each realisation unless otherwise stated. In each case, the final size is given as a proportion of the network size so that results for different values of n can be compared easily.

First, in Fig. 8 we compare the final size distribution for the approximate stochastic model for a homogeneous network with that for the full model. Results are shown for a network of $n = 1000$ nodes with $k = 50$ as the degree of each node, and rumour parameters $\lambda = 0.012$, $\delta = 0.1$, $p = 0.1$ so that $\psi = 1/6$. Note that, with $n = 1000$ and $p = 0.1$, the threshold value of ψ for the approximate model is $\psi = 0.92$ (see Fig. 3), so that the final size distribution is strongly bimodal. It can be seen that the two final size distributions are very similar. In the stochastic simulation, starting from a single spreader, the chance that the spreader 'forgets' the rumour before spreading it is $\delta/(\delta + \lambda k) = \psi/(1 + \psi)$: here this probability is $1/7$. Thus, in 1000 simulations, $1000/7 \simeq 143$ can be expected to die out immediately, with a standard deviation of $\sqrt{1000 \cdot 6/49} \simeq 11.1$, so that the observed value of 133 is not out of line.

One might expect that clustering of spreaders and stiflers in the simulations, not allowed for in the approximate model, would lead to slightly smaller final sizes, but any such effect is imperceptible here. Such clustering should also have an effect on the threshold for rumour spread. For the approximate model, if $\lambda = 0.012$, $k = 50$, the threshold value of $\psi = 1/6$ corresponds to a threshold value of $\delta = 0.547$. The effect of clustering should be that in the full model, the threshold value of δ would be lower (as dependence would result in less tendency of the rumour to spread, and thus less forgetting needed to control it). The limited simulations that have been conducted, varying δ and keeping other parameters fixed, indicate that this is the case, with a threshold for δ between 0.4 and 0.5.

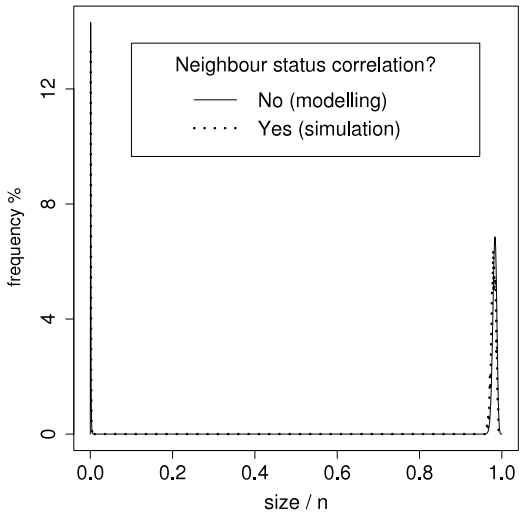
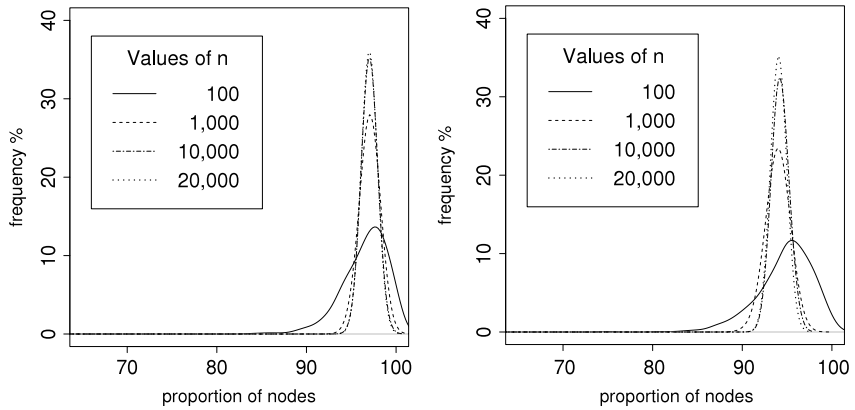
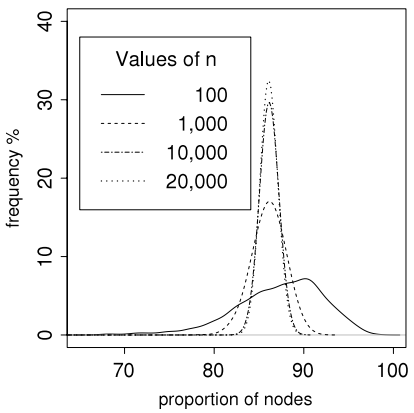


Fig. 8. Final size distribution comparing modelling without neighbour status correlation to simulation with it, $n = 1000$, $p = 0.1$, $\psi = 0.1667$.



(a) Homogeneous, $k = 10$.

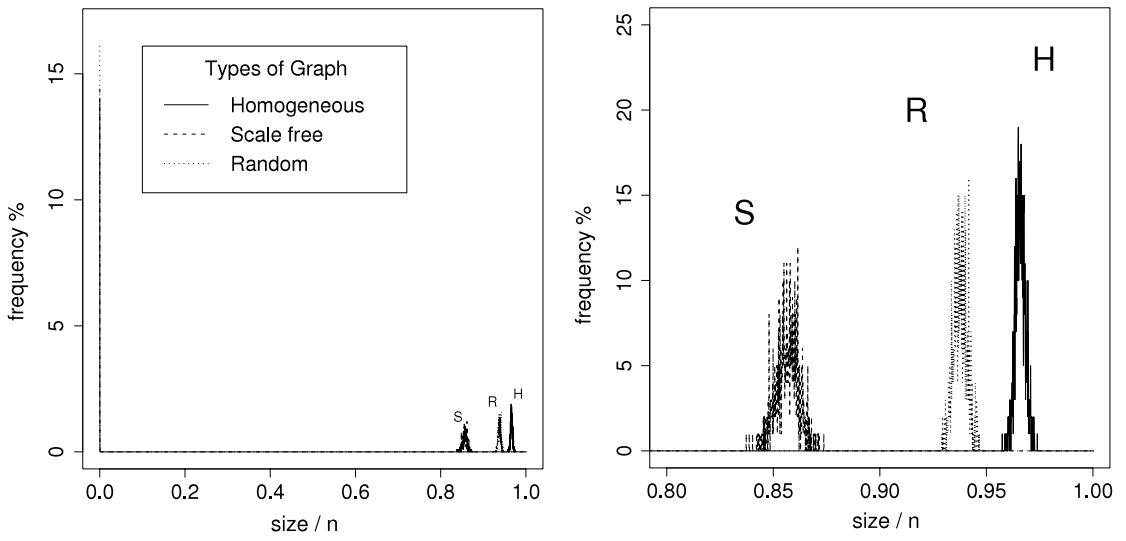
(b) Random, $\pi = 10/n$.



(c) Scale free, $\omega = 5$.

Fig. 9. Final size kernel densities for different types and sizes of networks, with $\lambda = 0.06$, $\delta = 0.1$ and $p = 0.1$.

To illustrate the effect of network size on the final size distribution for each of the three chosen network structures, some results are given in Fig. 9. For each structure, results are obtained for $n = 10^2$, 10^3, 10^4 and 2×10^4 (the limit given the computer time available). For the homogeneous graphs, $k = 10$, so that the graph has $5n$ edges. For the simple random graph the probability π was taken to be $10/n$, and for the scale-free option $\omega = 5$ edges were added at each stage in the



(a) Complete distribution. (b) Detail of second peak.

Fig. 10. Final size distribution for different networks but similar numbers of edges, parameter = 10/5/0.001, $n = 10{,}000$, $p = 0.1$, $\lambda = 0.06$, $\delta = 0.1$

construction, so that for each structure the expected number of edges is $5n$; the actual numbers of edges in the networks simulated was always very close to the mean. In order to compare distributions for a range of graph sizes, 100 equally sized bins were created for each size and kernel density plots drawn using common bandwidths. The first peaks are similar within and between plots and so are not shown in order to highlight the second peaks whose locations settle down once $n \geq 10^3$ (the similarity ensuring that the second peaks are comparable). As n increases, the scale-free graphs seem to exhibit most change in the height of the second peak with the homogeneous the least. This order reflects the variance in the degree of the vertices (vertices in homogeneous graphs all have the same degree so there is no variance while those in scale-free graphs are likely to vary considerably, as shown by the presence of hubs) and is described in more detail below.

In Fig. 10 the final size distributions are compared for the three network structures. The network sizes are taken to be $n = 10^4$ and parameters are chosen so that the mean number of edges is the same (50,000) in each of the three cases ($k = 10$ for the homogeneous graph, $\pi = 0.001$ for the simple random graph, $\omega = 5$ for the scale-free network). It can be seen that there is more 'distance' between the second peaks in the scale-free, and the homogeneous and random networks than between the latter two reflecting the fact that the rumour in the scale-free network, if it spreads, reaches a smaller proportion of the population. Note that in these simulations, the initial spreader is chosen at random and is therefore more likely to have a low degree in the scale-free case. Similarly, and on the basis of limited simulations, thresholds in δ (keeping all other parameters fixed) for rumour spread in the homogeneous and random networks seem similar while the threshold value of δ for the scale-free case is noticeably higher. This is consistent with the result of Ref. [3] that to first order in p, the threshold for δ is $\lambda(\mu_K + \sigma_K^2/\mu_K)$. Intuitively, the more variable the degree distribution of the nodes, the more control has to be exercised by the forgetting mechanism if the rumour is to be suppressed.

In all the results reported above, the rumour has started from an initial spreader chosen at random from the population, independently from realisation to realisation. The results therefore incorporate random variation due to the initial spreader as well as to the transmission of the rumour from that initial individual. The effect of the choice of initial spreader will be most noticeable for the scale-free networks. Thus, Fig. 11 seeks to disentangle these two sources of variation by showing the final size distribution for a scale-free network with 1000 nodes and 25,000 edges, for three cases: the initial node is the one with the maximum degree, the one with the minimum degree, or is chosen at random. No obvious difference between the three can be seen and to emphasise this, the figure just displays the detail of the second peak where we would expect any distinguishing features to be clearest.

Finally, we note that in the original description of the model, Ref. [3] assumed that spreaders broadcast rumours simultaneously to all neighbours, while we have assumed independent spreading to each neighbour separately. In a deterministic treatment there is no difference, but in a stochastic analysis there are considerable differences. We have assumed above that a spreader contacts each neighbour at rate λ so that, if the spreader has degree k, the total contact rate is λk. Thus we compare this situation with that in which the spreader makes broadcasts at rate λ, and where each broadcast goes to all neighbours.

Clearly, the rumour is much less likely to spread from the initial spreader in broadcast node (the probability that the rumour is forgotten before it is spread is $\delta/(\lambda + \delta)$, while in independent mode this probability is $\delta/(\lambda k + \delta)$). However, of more interest, as shown in the detail for the second peaks in Fig. 12, is that if the broadcast rumour does get transmitted by the initial spreader, and especially if the degree k is reasonably large, then the rumour will spread quickly to a larger

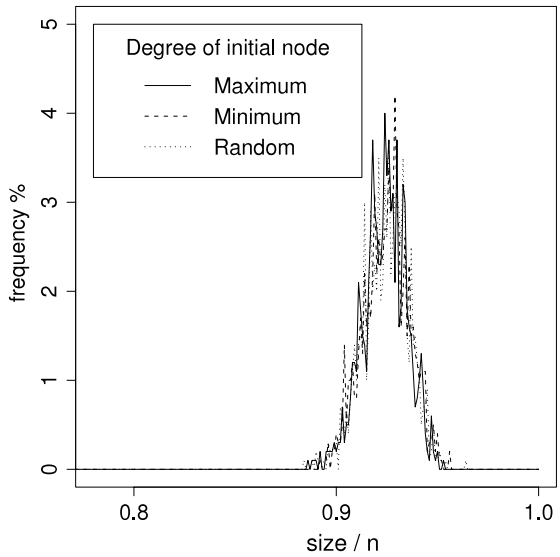
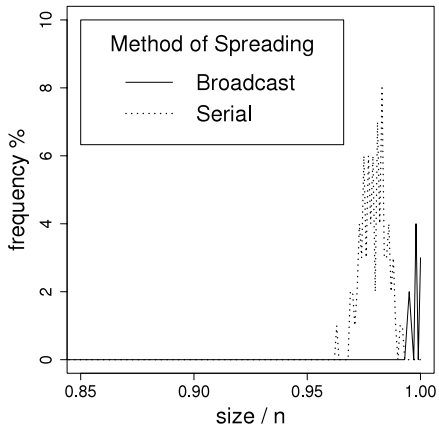
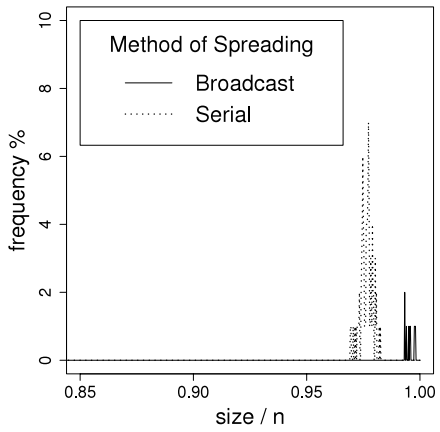


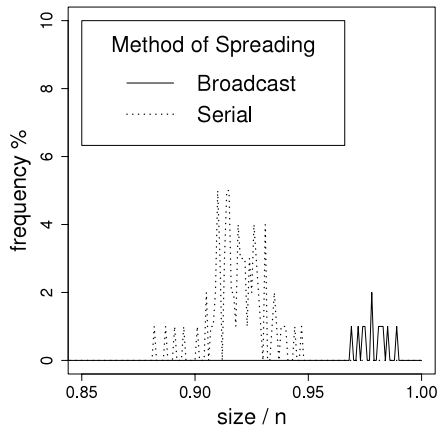
Fig. 11. Effect of choice of initial node for scale-free networks – maximum, minimum or random – details of second peak, $\omega = 25$, $n = 1000$, $p = 0.1$, $\lambda = 0.012$, $\delta = 0.1$.



(a) Homogeneous, $k = 50$.



(b) Random, $\pi = 0.0125$.



(c) Scale free, $\omega = 25$.

Fig. 12. Final size distribution for serial and broadcast networks of different types – details of second peak with $p = 0.1$, $\lambda = 0.012$, $\delta = 0.1$ and $n = 1000$ (for all but the random network where $n = 4000$ for ease of comparison).

proportion of the population than in the serial scenario. The former situation then more nearly resembles the independent case but with a large number of initial spreaders.

5. Discussion

In this paper we have investigated the stochastic spreading of epidemics and rumours on networks, focusing on the SIR epidemic model and a recently proposed rumour model on networks [3], and taking into account the structure of the underlying network at the level of the degree–degree correlation function. Using embedded Markov chain techniques and ignoring density correlations between neighbouring sites, we have derived a set of equations for the distribution of the final size of the epidemic/rumour on a homogeneous network. We then solved these equations numerically and compared the resulting distribution with the solution of the corresponding mean-field deterministic model. The final size distribution of the epidemic/rumour was found to switch from unimodal to bimodal form at a threshold value. For finite networks, this threshold value is higher than that obtained from the corresponding deterministic model. However, the difference between the two decreases with the network size, n, following a $n^{-1/3}$ behaviour. On the other hand, even in the limiting infinite system, the final size value obtained from the deterministic model does not coincide with the mean of the final size distribution from the stochastic model because of the bimodal structure of this distribution.

Next, the final size distribution for this model was compared with results from Monte Carlo simulations for the full stochastic model on a homogeneous network, where the latter incorporates both the complete network structure and the dependence of the transitions on the status of neighbouring nodes. It was found that the approximating stochastic model reproduces the exact simulation results with great accuracy. Finally, further Monte Carlo simulations were used to explore the effects of the network size and structure on the final size distribution (homogeneous networks, simple random graphs and the Barabasi–Albert scale-free networks were compared).

There has been significant recent research on stochastic spreading dynamics on networks. This large body of work has greatly increased our understanding of the role of network topology on both the threshold behaviour and dynamics of such processes. However, often analytical and numerical progress in incorporating the network structure was made at the expense of ignoring some of the important stochastic effects. In this paper we have taken a first step in incorporating and investigating such stochastic effects in the spreading of epidemics and rumours on networks. Our future work will focus on refining our approximate treatment of density correlations between neighbouring sites, and further numerical investigation of the stochastic spreading dynamics on networks that show strong clustering effects.

References

[1] M. Nekovee, Y. Moreno, Dynamics of gossip-like information dissemination in complex computer networks, J. Comp. Math. 85 (2008) 1165–1173.
[2] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, Adv. Physics 49 (2000) 815–958.
[3] M. Nekovee, Y. Moreno, G. Bianconi, M. Marsili, Theory of rumour spreading in complex networks, Physica A 374 (2007) 457–470.
[4] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[5] M. Boguna, R. Pastor-Satorras, A. Vespignani, Absence of epidemic threshold in scale-free networks with degree, Phys. Rev. Lett. 90 (2003) 028701.
[6] Y. Moreno, R. Pastor-Satorras, A. Vazquez, A. Vespignani, Critical load and congestion instabilities in scale-free networks, Europhys. Lett. 62 (2003) 292–298.
[7] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (2002) 521–529.
[8] I. Nåsell, The threshold concept in stochastic epidemic and endemic models, in: D. Mollison (Ed.), Epidemic Models: Their Structure and Relation to Data, Cambridge University Press, Cambridge, UK, 1995, pp. 71–83.
[9] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des advantages de l'inculation pour la prévenir, Mém. Math. Phys. Acad. Roy. Sci. Paris (1760) 1–45.
[10] W.H. Hamer, Epidemic disease in England, The Lancet 1 (1906) 733–739.
[11] R. Ross, The Prevention of Malaria, 2nd ed., Murray, London, UK, 1911.
[12] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics, Part I, Proc. R. Soc. Lond. A 115 (1927) 700–721.
[13] P. Whittle, The outcome of a stochastic epidemic - a note on Bailey's paper, Biometrika 42 (1955) 116–122.
[14] P. Whittle, On the use of the normal approximation in the treatment of stochastic processes, J. Roy. Statist. Soc. B 19 (1957) 268–281.
[15] D.J. Daley, D.G. Kendall, Stochastic rumours, J. Inst. Math. Appl. 1 (1965) 42–55.
[16] T.G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Probab. 7 (1970) 49–58.
[17] T.G. Kurtz, Limit theorems for sequences of jump Markov processes approximating ordinary differential processes, J. Appl. Probab. 8 (1971) 344–356.
[18] T.G. Kurtz, Approximation of Population Processes, SIAM, Philadelphia, USA, 1981.
[19] A.D. Barbour, The principle of diffusion of arbitrary constants, J. Appl. Probab. 9 (1972) 519–541.
[20] A.D. Barbour, On a functional central limit theorem for Markov population processes, Adv. Appl. Prob. 6 (1974) 21–39.
[21] H.E. Daniels, A look at perturbation approximations for epidemics, in: I.V. Basawa, R.L. Taylor (Eds.), Selected Proc. Sheffield Symp. Appl. Probab, IMS Monograph Series, vol. 18, Hayward, California, USA, pp. 48–65.
[22] A.D. Barbour, The duration of the closed stochastic epidemic, Biometrika 62 (1975) 477–482.
[23] D. Mollison, Dependence of epidemic and population velocities on basic parameters, Math. Biosci. 107 (1991) 255–287.
[24] D.J. Daley, J. Gani, Epidemic Modelling, Cambridge University Press, Cambridge, UK, 1999.
[25] D.G. Kendall, La propagation d'une épidémie ou d'un bruit dans une population limitée, Publ. Inst. Statist. Univ. Paris 6 (1957) 307–311.
[26] D.J. Daley, D.G. Kendall, Epidemics and rumours, Nature 204 (1964) 1118.
[27] D.P. Maki, M. Thompson, Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences, Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1973.
[28] Y. Moreno, M. Nekovee, A.F. Pacheco, Dynamics of rumour spreading in complex networks, Phys. Rev. E 69 (2004) 066130/1-066130/7.
[29] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[30] Daniel T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (1977) 2340–2361.
[31] F. Ball, D. Mollison, G. Scalia-Tomba, Epidemics with two levels of mixing, Ann. Appl. Probab. 7 (1997) 46–89.
[32] B. Bollobás, Random Graphs, Cambridge University Press, Cambridge, UK, 2001.
[33] M.E.J. Newman, Properties of highly clustered networks, Phys. Rev. E 68 (2003) 026121.

[34] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Structures Algorithms 6 (1995) 161–179.
[35] H.A. Simon, On a class of skew distribution functions, Biometrika 42 (1955) 425–440.
[36] A-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[37] G. Csanyi, B. Szendroi, Structure of large social networks, Phys. Rev. E 69 (2004) 036131/1-036131/5.
[38] F. Ball, I. Nåsell, The shape of the size distribution of an epidemic in a finite population, Math. Biosci. 123 (1994) 167–181.