Math. Scientist **24**, 128–147 (1999)
Printed in England

EPIDEMIC MODELS AND SOCIAL NETWORKS

HÅKAN ANDERSSON,* *Stockholm University*

Abstract

This survey paper discusses a class of stochastic continuous time models for the spread of an epidemic across a static or dynamic social network. Various simple graphs are considered: Bernoulli random graphs, graphs with prescribed degrees, graphs with a certain number of short loops, overlapping subgraphs representing the superposition of independent networks, and dynamically changing graphs. For each of these, expressions for important epidemiological quantities such as the basic reproduction number, the final size of the epidemic and the time dynamics of the proportion of susceptible and infectious individuals, are derived. The modelling assumptions are meaningful for finite populations, but the results obtained are only valid asymptotically as the population size tends to infinity. The theoretical work is illustrated by computer simulations and numerical calculations.

Keywords: Epidemic processes; final epidemic size; basic reproduction number; random graphs

AMS 1991 Subject Classification: Primary 92D30
Secondary 05C80; 60J27

1. Introduction

There is a considerable body of literature concerned with mathematical models of the spread of infectious diseases. The book by Bailey [12] provides an excellent introduction to the theory of epidemic models, while a comprehensive treatment of the subject is given by Anderson and May [6]. The vast majority of models for epidemic spread studied so far are much too simplified to be of any practical interest; it is indeed argued that such toy models have the sole purpose of providing an amusement for the theoretician! It is of course impossible to set strictly realistic modelling assumptions, since even minor modifications of the basic models often lead to intractable mathematics. It remains extremely important, however, to continue developing stochastic and deterministic epidemic models, along with the mathematical tools to analyse them. In the future, we may gain enough knowledge to be able to predict with some accuracy the important characteristics of a real life epidemic. Needless to say, any such knowledge would be invaluable from a public health point of view.

We present a survey of a class of theoretical epidemic models in which the assumption of *homogeneous mixing* is relaxed, while in all other respects the simplest possible setup is chosen. Some new ideas and results are also discussed, in particular in Sections 4 and 6. We first give a simple argument pointing out a serious drawback in the assumption of homogeneous mixing. The stochastic epidemic model that has received the most attention is the so-called *standard epidemic process* (following the terminology of Watson [74]). The model is better known as the general stochastic epidemic, but this term, which originated with Bailey [11], is

Received 14 October 1998.
* Postal address: Group Financial Risk Control, Swedbank, S-10534 Stockholm, Sweden.
Email address: hakan.b.andersson@foreningssparbanken.se

:4, 128–147 (1999)
Printed in England
bability Trust 1999

now inappropriate since the model has been greatly generalized over the years. This model can be described as follows. Consider a closed population consisting of n individuals. Initially there are a infectious and $n - a$ susceptible individuals. An individual remains infectious for an exponential holding time with intensity 1, during which it makes contact with each susceptible at the time points of a time homogeneous Poisson process with intensity β/n. Such a contact results in the susceptible individual becoming infectious. An individual is considered 'removed' once its infectious period has terminated, and is then immune to new infections, playing no further part in the epidemic.

Many of the classical results for the standard epidemic process (branching approximations, deterministic approximations, calculations of the basic reproduction number and of the probability of a large outbreak, etc.) require the population size n to be *large*. But the modelling assumptions specify that in a large population, contacts between two given individuals occur at a very low rate, thus implying in principle that the possibility of repeated contacts is not taken into account. The assumption of homogeneous mixing is therefore not very realistic when describing epidemic spread in large populations. Moreover, as we will see later (notably in Section 3), even the slightest departure from homogeneous mixing may very well lead to dramatic changes in the observable epidemiological quantities.

If we wish to allow repeated contacts between individuals, a possible solution is to pick a graph describing the relations between individuals, and then let the disease spread along the social network so obtained. In that way each individual will be assigned a small neighbourhood of other individuals and can then contact each of its neighbours at a 'normal' rate. It is far from obvious how to choose a suitable form of the network. The graph should be complicated enough to catch something of the often irregular contact pattern in a population of living organisms, but at the same time simple enough to lend itself to mathematical analysis. In this work a number of simple graphs are proposed: Bernoulli random graphs, graphs with prescribed degrees, graphs with a certain number of short loops, overlapping subgraphs representing the superposition of independent networks, and graphs that change dynamically with time (Sections 3–7, respectively). The spread of disease is then modelled on each of these structures. We are particularly interested in the basic reproduction number, the asymptotic final size of the epidemic and the time dynamics of the asymptotic proportion of susceptible and infectious individuals. No proofs are given; the reader will find these in the references. The theory is illustrated by computer simulations and numerical calculations.

2. Preliminaries

We begin by setting up the modelling assumptions used in most of the work; in Section 7 it will be necessary to make a slight modification of these assumptions. Definitions of the basic reproduction number and the final epidemic size will also be given. Finally, we indicate how some calculations may be simplified by studying the progress of the disease on a generation basis rather than in real time.

2.1. Standard epidemic process on a fixed graph

Consider a closed population consisting of n individuals. Represent the neighbourhood structure in the population with a labelled undirected graph $\mathcal{G}$, so that the ith and the jth vertices of the graph are connected by an edge if and only if individuals i and j are neighbours. The graph will often be the result of some random experiment. We assume that the structure is fixed during the course of the epidemic; this is a reasonable simplification when considering short-term epidemics, since a typical social network is not expected to vary too much over a

short time span. Then let G be the adjacency matrix of the graph $\mathcal{G}$, so that $G_{ij} = 1$ if i and j are connected and $G_{ij} = 0$ otherwise. It follows that G is a symmetric binary $n \times n$ matrix with zeros in the diagonal. Finally, the degree D_i of the ith vertex is defined as the number of vertices adjacent to this vertex, $D_i = \sum_{j=1}^{n} G_{ij}$.

Let us next define the dynamics of the epidemic process. We pick a initially infectious individuals at random from the population. An infectious individual remains so for an exponentially distributed time period of intensity 1. During this period it makes close contacts with each of its neighbours according to the points of a Poisson process with intensity β. If the individual so contacted is still susceptible, then it will immediately become infectious. After the infectious period, the infectious individual recovers and is then immune. All infectious periods and Poisson processes are assumed to be independent of each other.

2.2. Important epidemiological quantities

We are particularly interested in the behaviour of the process as the population size n tends to infinity. Thus we are actually considering a *sequence* of epidemic processes indexed by n. We shall however be somewhat careless in our notation, suppressing this index throughout the work. Let us first define the final size of the epidemic. Let $X_i(t) = 1$ if the ith individual is susceptible at time t, and $X_i(t) = 0$ otherwise ($1 \leq i \leq n$). Likewise, $Y_i(t) = 1$ if the ith individual is infectious at time t, and $Y_i(t) = 0$ otherwise. Then define

$$X(t) = \sum_{i=1}^{n} X_i(t) \quad \text{and} \quad Y(t) = \sum_{i=1}^{n} Y_i(t)$$

to be the total number of susceptibles and infectives, respectively, at time t. Also, let T be the random time when the epidemic terminates, i.e.

$$T = \inf\{t \geq 0 : Y(t) = 0\}.$$

Then the number of individuals that *escape* the epidemic is given by $X(T)$. We define the final size of the epidemic to be the random number $n - X(T)$. Note that according to this definition, the initially infectious individuals are included in the final size.

The basic reproduction number R_0 has a long and interesting history, see for example [32], [33] and [43]. This quantity is traditionally defined as the average number of secondary cases generated by one infectious individual in a large completely susceptible population. However, when studying multitype populations or populations where there are repeated contacts between individuals, it soon becomes clear that this simple definition must be modified to provide meaningful information about the progress of the disease (see [29]). Here, we shall simply define R_0 directly in terms of the epidemic behaviour that the basic reproduction number is meant to reflect.

Assume that the initial number of infectious individuals, a, is kept *fixed* rather than growing with n. We say that a major epidemic (or a large outbreak) occurs if the final size is of the same order of magnitude as the population itself. Otherwise, if the final size is $o(n)$ only a minor epidemic occurs. Now, R_0 is defined as *any* non-negative function of the infection rate β and some graph characteristics such that if $R_0 \leq 1$ then the asymptotic probability of a major epidemic is zero, while if $R_0 > 1$ then there is, asymptotically, a strictly positive probability of a large outbreak. With this definition, R_0 is far from unique; if ϕ is any non-negative strictly increasing function such that $\phi(1) = 1$, then $\phi(R_0)$ is also a basic reproduction number. Nevertheless, whenever a very natural candidate for R_0 has been derived, we shall refer to it as *the* basic reproduction number.

Suppose now that the initial proportion a/n of infectious individuals converges to a number $\bar{a}$ strictly between 0 and 1 as n tends to infinity. As above, denote 'susceptibles' and 'infectives' by the letters X and Y, respectively. Generally speaking we set $A_i = 1$ if the ith individual is of type A, and $A_i = 0$ otherwise. Then define the scaled number of individuals, connected pairs and connected triples with given type configurations in the following way:

$$a = \frac{1}{n}\sum_{i=1}^{n} A_i,$$

$$[ab] = \frac{1}{n}\sum_{i,j=1}^{n} A_i G_{ij} B_j,$$

$$[abc] = \frac{1}{n}\sum_{\substack{i,j,k=1 \\ i\neq k}}^{n} A_i G_{ij} B_j G_{jk} C_k. \tag{1}$$

For instance, $[xy](t)$ is the (scaled) number of neighbours i, j where i is susceptible and j is infectious at time t. If the total number of neighbours of a given individual is kept bounded, then all sums above are $O(n)$, which explains the scaling by n. Note that $[ab] = [ba]$, and that each pair in $[aa]$ is counted twice. The variables counting the number of pairs and triples of different types will turn out to be most useful in the model analysis to come. Finally, define σ to be the proportion of individuals that escape the epidemic, and denote by τ the proportion of individuals that have ever experienced the disease,

$$\sigma = \frac{X(T)}{n} \quad \text{and} \quad \tau = 1 - \frac{X(T)}{n}. \tag{2}$$

Since $0 < \bar{a} < 1$, we expect from the law of large numbers that the quantities in Equations (1) and (2) will all have well-defined deterministic limits as n tends to infinity.

2.3. Generation process

The study of the basic reproduction number as well as the final epidemic size is sometimes greatly simplified by considering the epidemic process on a generation basis, see for example Ludwig [60] and von Bahr and Martin-Löf [10]. Let $\mathcal{X}(0)$ be the set of initially susceptible individuals, and denote by $\mathcal{Y}(0)$ the set of initial infectives,

$$\mathcal{X}(0) = \{i : X_i(0) = 1\} \quad \text{and} \quad \mathcal{Y}(0) = \{i : Y_i(0) = 1\}.$$

Then, for each $k \geq 1$, let the set $\mathcal{X}(k)$ consist of those in $\mathcal{X}(k-1)$ who were not infected by the members of $\mathcal{Y}(k-1)$, and put $\mathcal{Y}(k) = \mathcal{X}(k-1) \setminus \mathcal{X}(k)$. In this way a discrete time process $(\mathcal{X}(t), \mathcal{Y}(t))$; $t = 0, 1, 2, \ldots$, called the generation process, is obtained. We destroy the time order of events by this construction, but note that the directed transmission links are exactly the same in the generation process as in the original one. Hence observables that depend only on the size of the epidemic and not on any real time behaviour may be calculated by using the discrete time process. For instance, consider the basic reproduction number. If the population is large and the initial number of infectives is kept constant then the process $|\mathcal{Y}(t)|$; $t = 0, 1, 2, \ldots$, counting the number of infectives, often follows some kind of branching behaviour. Classical branching process theory may then be invoked to find conditions for possible explosion, i.e. to find R_0.

3. Bernoulli random graphs

The so-called Bernoulli random graph model is probably the simplest non-trivial choice of underlying network for our epidemic model. The theory of random graphs was introduced by Erdös and Rényi [37] in the late 1950s and has been extensively studied ever since. In particular, the class of Bernoulli random graphs has been very thoroughly explored, see for example [25] and references therein. This graph model, often referred to simply as the $\mathcal{G}(n, p)$ model, is defined as follows. We are given a set of n labelled vertices. With probability p, we connect a given pair of distinct vertices i and j by drawing an edge between them. These connections are made independently of each other. It is clear that the degree D_i of the ith vertex, i.e. the number of vertices adjacent to this vertex, is binomially distributed with parameters $n-1$ and p, so that the average degree is given by $(n-1)p$. In order to keep the size of the neighbourhood bounded we have to set $p = \lambda/n$, for some $\lambda > 0$. This implies that D_i will be approximately Poisson distributed with parameter λ if n is large. Now, for each outcome $\mathcal{G}$ belonging to $\mathcal{G}(n, p)$ we run the standard epidemic on $\mathcal{G}$, as described in Section 2. This model is described also in [69]. Related papers in random graph theory are for example [72], [46] and [45]. See also the book on random mappings by Kolchin [52]. The corresponding generation process is covered for example in [61] and [68]; see also [22], [23], [40] and [15].

3.1. Basic reproduction number

First, let us calculate the basic reproduction number. Denote by p_k the Poisson probability of having k neighbours, $k \geq 0$. Consider the generation process of infectious individuals, starting with one single infective. In the beginning of the time course, all contacted individuals are susceptible with high probability since the population is assumed to be large. Hence the generation process of infectives is well approximated by a branching process. The reproduction mean of this branching process will serve as our basic reproduction number since, according to classical branching process theory (see for example [42]), the process has a positive probability of exploding if and only if this mean exceeds 1. One should consider an infective in the second generation; the initial infective may be atypical, since *all* its neighbours are susceptible. An infective in the second generation will have k neighbours with a probability proportional to kp_k, and in this case generates on the average $(k-1)\beta/(\beta+1)$ new cases. Thus

$$R_0 = \frac{\beta}{\beta+1} \sum_{k=1}^{\infty} (k-1) \cdot \frac{kp_k}{\sum_{k'=1}^{\infty} k' p_{k'}},$$

which can be rewritten as

$$R_0 = \frac{\beta}{\beta+1} \left(\mathrm{E}(D) + \frac{\mathrm{var}(D) - \mathrm{E}(D)}{\mathrm{E}(D)} \right). \tag{3}$$

For this particular model, $\mathrm{E}(D) = \mathrm{var}(D) = \lambda$, implying that

$$R_0 = \frac{\lambda\beta}{\beta+1}. \tag{4}$$

3.2. Time dynamics

We now describe the time dynamics of the model. Assume that the proportion of initially infectious individuals tends to a nontrivial limit $\bar{a}$ as n tends to infinity. The quantities introduced in Equation (1) turn out to be of fundamental importance here. In [69] the following system of differential equations, which is valid asymptotically as the population size tends to infinity, is derived:

$$\begin{aligned}
\frac{dx}{dt} &= -\beta[xy], \\
\frac{dy}{dt} &= \beta[xy] - y, \\
\frac{d[xx]}{dt} &= -2\beta[xxy], \\
\frac{d[xy]}{dt} &= \beta([xxy] - [yxy] - [xy]) - [xy], \\
\frac{d[yy]}{dt} &= 2\beta([yxy] + [xy]) - 2[yy].
\end{aligned} \tag{5}$$

We proceed to explain the fourth line; the other lines are then obtained by similar reasoning. If the central individual j in a connected XXY-triple (i, j, k) is infected by the individual k, we gain an XY-pair (i, j). On the other hand, we may lose an XY-pair (j, k) in three ways: The individual j in an YXY-triple (i, j, k) is infected by the individual i; k infects j directly; k becomes removed. The fourth line of (5) now follows readily.

The system (5) is not very useful as it stands, since no description of the time dynamics of the variables $[abc]$ is provided. Fortunately, the equations can be closed at the level of pairs by the following device. Consider a connected triple (i, j, k). Then, using the notation introduced in Section 2,

$$\mathrm{P}(A_i B_j C_k = 1) = \mathrm{P}(A_i = 1 \mid B_j C_k = 1)\,\mathrm{P}(B_j C_k = 1).$$

If n is large then the probability that (i, j, k) is part of a short closed loop (e.g. a triangle) is negligible, hence asymptotically,

$$\begin{aligned}
\mathrm{P}(A_i = 1 \mid B_j C_k = 1) &= \mathrm{P}(A_i = 1 \mid B_j = 1) \\
&= \frac{\mathrm{P}(A_i B_j = 1)}{\mathrm{P}(B_j = 1)}.
\end{aligned}$$

This translates to the formula

$$[abc] = \frac{[ab][bc]}{b} \qquad \text{for all } a, b, c. \tag{6}$$

Finally, we insert this relation into the system (5) to obtain a closed system in the five variables x, y, $[xx]$, $[xy]$ and $[yy]$.

This system can be simplified considerably. Obviously, the equation for $[yy]$ is superfluous. Also the equation for $[xx]$ can be crossed out, since by comparing the differential equations for x^2 and $[xx]$ it follows that $[xx] = \lambda x^2$ at all times. Finally, we define $\hat{y} = [xy]/x$. This quantity can be viewed as the *infection pressure* exerted on the population. It is easily verified that

$$\frac{d\hat{y}}{dt} = \frac{1}{x^2}\left(x\frac{d[xy]}{dt} - [xy]\frac{dx}{dt}\right) = (\lambda\beta x - \beta - 1)\,\hat{y},$$

giving the system of differential equations

$$\begin{aligned} \frac{dx}{dt} &= -\beta x \hat{y}, \\ \frac{dy}{dt} &= \beta x \hat{y} - y, \\ \frac{d\hat{y}}{dt} &= (\lambda\beta x - \beta - 1)\,\hat{y}, \end{aligned} \tag{7}$$

with initial condition $x(0) = 1 - \bar{a}$, $y(0) = \bar{a}$ and $\hat{y}(0) = \lambda\bar{a}$.

Note that if $\lambda \to \infty$ and $\beta \to 0$ in such a way that their product $\lambda\beta$ is kept constant, then in the limit $\hat{y} = \lambda y$ (by applying Gronwall's inequality to the function $\hat{y} - \lambda y$). Thus (7) collapses to the two-dimensional system

$$\begin{aligned} \frac{dx}{dt} &= -\lambda\beta x y, \\ \frac{dy}{dt} &= \lambda\beta x y - y, \end{aligned} \tag{8}$$

$x(0) = 1 - \bar{a}$, $y(0) = \bar{a}$. We recognize in (8) the usual Kermack–McKendrick model (the deterministic version of the classical standard epidemic process), see Bailey [12].

3.3. Final size equation

Remember that the final size τ is given by $\tau = 1 - \sigma$, where σ is the relative number of individuals that escape the epidemic. An equation for this latter quantity is easily derived. Divide the third line of (7) by the first one and integrate to get

$$\frac{\beta + 1}{\beta} \log\left(\frac{x}{1 - \bar{a}}\right) = -\lambda(1 - x) + \hat{y}.$$

Putting $\hat{y} = 0$ yields

$$\sigma = (1 - \bar{a}) \exp\left\{-\frac{\lambda\beta}{\beta + 1}(1 - \sigma)\right\}. \tag{9}$$

Note that this transcendental equation is of the same form as the final size equation corresponding to the Kermack–McKendrick model. The basic reproduction number R_0 from (4) appears in the formula (9) as it should.

3.4. Numerical results

Let us point out an interesting property of the present model. From (7) we observe that in the limiting case $\bar{a} \to 0$ the infection pressure $\hat{y}$ is increasing initially if and only if $R_0 > 1$. On the other hand, the same is *not* true for the proportion of infectious individuals. Indeed,

$$\frac{dy}{dt} = \beta x \hat{y} - y = \left(\frac{\beta x \hat{y}}{y} - 1\right) y,$$

thus y is increasing initially if and only if $\lambda\beta > 1$ (in the limit $\bar{a} \to 0$). Hence there exists a region in the (β, λ)-parameter space, given by

$$\frac{\lambda\beta}{\beta + 1} < 1 \qquad \text{and} \qquad \lambda\beta > 1,$$

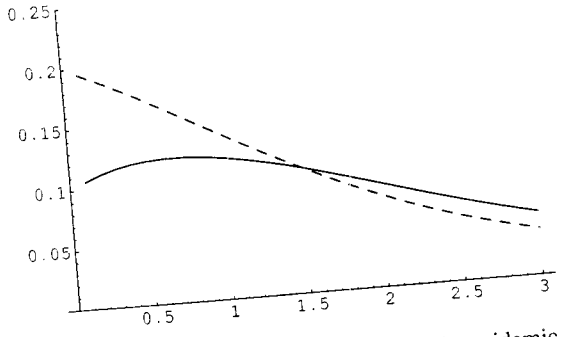

FIGURE 1: Deceptive initial behaviour of the epidemic.

where the epidemic first seems to take off but soon flattens out and declines. This phenomenon is illustrated in Figure 1, where the functions $y(t)$ (solid line) and $\hat{y}(t)$ (dashed line) are plotted against t for the case $\lambda = 2$, $\beta = 0.95$ and $\bar{a} = 0.1$. To conclude, we stress that it is the initial behaviour of the infection pressure rather than the proportion of infectious individuals that determines whether or not a large outbreak occurs.

4. Random graphs with prescribed degrees

In this section we study a more general class of networks. Consider the usual closed population consisting of n individuals, and suppose that we have gained information about how the neighbourhood size is distributed, but otherwise have no information on the network. Then it is natural to first pick a sample of n numbers according to this distribution and then choose a labelled graph having exactly these numbers as degrees. This graph will provide a very crude model of the social network.

More formally, for each n define a vector $(D_1, \ldots, D_n)$ of identically distributed random variables to represent the degrees of the graph. We think of the variables D_i as independent, but this is of course not strictly true; for instance, the sum $\sum_{i=1}^{n} D_i$ should always be an even number. Also, assume that D_i tends to a proper random variable D in distribution as $n \to \infty$. Finally, among all labelled undirected graphs on n vertices with given degrees D_i, pick one uniformly at random and run the standard epidemic on this structure. Unless the distribution of D has a very heavy tail, this random experiment will produce graphs with a high probability of containing very few loops (see [8] for convenient technical conditions). This is the reason why the deterministic approximation presented below works so well. On the other hand, from the point of view of applications, we would of course appreciate the presence of short loops such as triangles ('two friends often have common friends'). In Section 5, a method of incorporating short loops is suggested. In [8] a simple stochastic discrete time epidemic process on the above network is discussed; see also [7]. Diekmann *et al.* [28] also study the special case of a regular graph, i.e., a graph where each vertex has the same degree. Among relevant papers from the graph theoretic field are [21], [63] and [66].

4.1. Basic reproduction number

The derivation of the basic reproduction number given in Section 3 is still valid:

$$R_0 = \frac{\beta}{\beta + 1}\left(\mathrm{E}(D) + \frac{\mathrm{var}(D) - \mathrm{E}(D)}{\mathrm{E}(D)}\right). \tag{10}$$

Note that even if the average neighbourhood size is small, the presence of some individuals with very large circles of acquaintance may very well lead to a dangerously high basic reproduction number.

4.2. Final size equation

The proportion σ of individuals that escape the epidemic can be derived as follows. First, $\sigma = (1-\bar{a})\sum_{k=0}^{\infty}\xi_k p_k$, where ξ_k is the chance that an initially susceptible individual with in total k neighbours stays susceptible throughout the epidemic. If we define s to be the probability that a *neighbour* of such an individual escapes the epidemic, then

$$\xi_k = \sum_{\ell=0}^{k}\left(\frac{1}{\beta+1}\right)^{\ell}\binom{k}{\ell}s^{k-\ell}(1-s)^{\ell},$$

where $1/(\beta+1)$ is the probability of not receiving the disease from a given infectious neighbour. Thus

$$\xi_k = \left(\frac{1+\beta s}{1+\beta}\right)^{k},$$

and it follows that

$$\sigma = (1-\bar{a})\varphi\left(\frac{1+\beta s}{1+\beta}\right), \tag{11}$$

where φ is the probability generating function of D. By similar reasoning we derive an equation for s, the probability that a given neighbour escapes the epidemic. We have $s = (1-\bar{a})\sum_{k=0}^{\infty}\tilde{\xi}_k\tilde{p}_k$, where $\tilde{p}_k$ is the probability that our neighbour has in turn k neighbours, and $\tilde{\xi}_k$ denotes the probability that he or she stays susceptible given k neighbours in total. We easily obtain

$$\tilde{p}_k = kp_k \Big/ \left(\sum_{k'=0}^{\infty}k' p_{k'}\right),$$

$$\tilde{\xi}_k = \sum_{\ell=0}^{k-1}\left(\frac{1}{\beta+1}\right)^{\ell}\binom{k-1}{\ell}s^{k-1-\ell}(1-s)^{\ell} = \left(\frac{1+\beta s}{1+\beta}\right)^{k-1},$$

implying that

$$s = (1-\bar{a})\varphi'\left(\frac{1+\beta s}{1+\beta}\right)\Big/\varphi'(1). \tag{12}$$

Equations (11) and (12) together determine the escaping proportion σ. It is clear that the solution is unique, since the right-hand side minus the left-hand side of (12) defines a convex function of s which is strictly positive at $s = 0$ and strictly negative at $s = 1$.

As a simple example, assume that the average neighbourhood size D is Poisson distributed with parameter λ; this is the case considered in the last section. Equation (4) yields $R_0 = \lambda\beta/(\beta+1)$. The probability generating function is given by $\varphi(z) = \exp\{-\lambda(1-z)\}$. It is readily seen that $s = \sigma$ and then

$$\sigma = (1-\bar{a})\exp\left\{-\frac{\lambda\beta}{1+\beta}(1-\sigma)\right\},$$

in accordance with the results of the last section. Another natural example is to let D be constant, $D \equiv k$. This is the case treated in [28]. By (10), $R_0 = (k-1)\beta/(\beta+1)$ since $\mathrm{var}(D) = 0$. To calculate the escaping proportion σ we note that $\varphi(z) = z^k$, implying that

$$\sigma = (1-\bar{a})\left(\frac{1+\beta s}{1+\beta}\right)^k \quad \text{where } s = (1-\bar{a})\left(\frac{1+\beta s}{1+\beta}\right)^{k-1}.$$

These equations can be turned into a single equation for σ, but unless $k = 1$ or $k = 2$ they cannot be solved explicitly.

4.3. Simulations and numerical results

We now use simulations to investigate the validity of the formulas above. In order to generate a labelled graph with n vertices and fixed degrees $(D_1, \ldots, D_n)$ we proceed as follows:

- Form a list $\mathcal{L}$ containing D_i copies of the ith vertex, $1 \le i \le n$;
- Choose a random pairing of the elements of $\mathcal{L}$;
- Repeat the above step as long as there exist pairs (i, i) or else different pairs involving the same two vertices i and j.

When this algorithm has been completed, for each pair (i, j) of the pairing we draw an edge to connect the ith and the jth vertex. In this way a graph with the correct degrees is obtained. Also, since any of these graphs can be obtained in exactly $\prod_{i=1}^{n} D_i!$ ways, it follows that all the graphs are equally probable as desired. Having generated such a graph, it is a simple task to simulate the standard epidemic.

Some results for the final size $\tau = 1 - \sigma$ are shown in Table 1. Using two different values of the population size n ($n = 1000$ and $n = 10\,000$) we have estimated the average and the standard deviation of the random final size τ_{sim}. This should be compared with the value τ_{det} obtained from (11) and (12). Here $D_1 \equiv 4$, D_2 is binomially distributed with parameters 2 and $\frac{1}{2}$, translated to get the expected value 4, and D_3 is binomial with parameters 8 and $\frac{1}{2}$. Finally, D_4 has distribution given by $p_2 = p_6 = \frac{1}{2}$; $p_k = 0$ otherwise. In this way the degree variables have the same average but increasing variance.

The discrete time epidemic model of [8] exhibits interesting behaviour. The model is defined on exactly the same type of network as the one considered in this section, and it turns out that for a highly infectious disease, a network with a constant number of neighbours yields a larger final size than any irregular network with the same average neighbourhood size. On the other hand, if the disease is less contagious, the regular network is *not* extreme in this respect. Further examples related to this observation are presented in [9]. The phenomenon is also present for our continuous time process, and is illustrated in Figure 2 where the final size is plotted against β for two different choices of the degree variable D (we put $\bar{a} = 0.1$). The solid line corresponds to $p_3 = 1$; $p_k = 0$ otherwise, and the dashed line is obtained by using a network with $p_2 = p_4 = \frac{1}{2}$; $p_k = 0$ otherwise.

5. Random graphs containing short loops

As mentioned in the last section, there are always several small groups with complete mixing in a typical social network. It would therefore be interesting to find a natural way of constructing random graphs that contain short loops. In [39] the class of so-called *Markov*

TABLE 1: Final size of epidemic for various networks.

		τ_{sim}; $n = 1000$		τ_{sim}; $n = 10\,000$		
D	β	Average	s.d.	Average	s.d.	τ_{det}
D_1	0.1	0.142	0.016	0.142	0.005	0.142
	0.5	0.519	0.051	0.523	0.016	0.524
	1.0	0.885	0.020	0.887	0.007	0.887
D_2	0.1	0.144	0.015	0.143	0.005	0.143
	0.5	0.531	0.049	0.536	0.016	0.537
	1.0	0.878	0.021	0.880	0.007	0.880
D_3	0.1	0.145	0.016	0.145	0.005	0.145
	0.5	0.565	0.044	0.569	0.014	0.569
	1.0	0.855	0.019	0.857	0.006	0.856
D_4	0.1	0.146	0.017	0.145	0.006	0.147
	0.5	0.590	0.038	0.592	0.012	0.593
	1.0	0.823	0.019	0.823	0.006	0.824

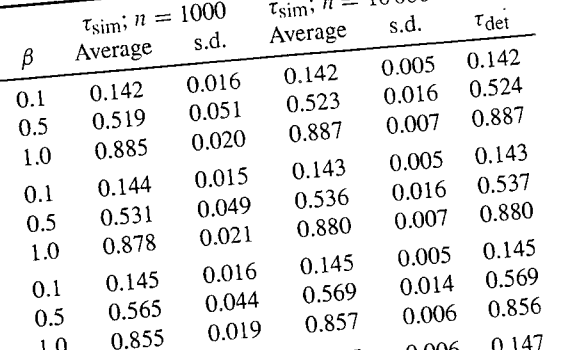

FIGURE 2: Trade-off in the final size.

graphs is considered. In the simplest case a given undirected labelled graph $\mathcal{G}$ on n vertices has probability

$$C \exp\left\{\sum_{j=1}^{n-1} \delta_j Z_j + \gamma T\right\},$$

where Z_j is the number of vertices of degree j and T is the number of triangles in $\mathcal{G}$. Positive (negative) values of the parameters δ_j indicate a tendency towards (against) many vertices of degree j. The parameter γ similarly controls the number of triangles. At first sight this seems to be a promising construction, but it is flawed. Strauss [73] shows that the model is degenerate in the sense that if γ is strictly positive (however small) then, as the number of vertices increases, the probability that an arbitrarily large fraction of the edges will coalesce into a clique, i.e. a complete subset of the graph, tends to 1. Jonasson [47] tries to improve the Markov graph model by letting γ depend on the number of vertices, but it turns out that the resulting model is explosive. When n is large, typical outcomes are *either* almost free from triangles *or* almost complete. These results together with simulations by Altmann [3] show

that Markov graphs have undesirable asymptotic properties. The construction proposed below produces graphs that are rather well-behaved, but to find even more natural alternative models remains an open problem.

Following the construction of Section 4, for each n we let $D_1, \ldots, D_n$ be identically distributed random variables, and for each sample $(D_1, \ldots, D_n)$ we define P_0 to be the uniform measure on the labelled undirected graphs with n vertices and with exactly these numbers as degrees. Then, for a given outcome $\mathcal{G}$ let $\Phi(\mathcal{G})$ be the ratio of the number of triangles in $\mathcal{G}$ to the number of connected triples in $\mathcal{G}$. Finally, for given $\theta \geq 0$ and ϕ, $0 \leq \phi \leq 1$, we define the measure

$$P_\theta^\phi(\mathcal{G}) = C P_0(\mathcal{G}) \exp\left\{-\theta n |\Phi(\mathcal{G}) - \phi|\right\}.$$

Note that $P_0^\phi = P_0$, but by choosing larger θ we increasingly reward outcomes having the ratio of triangles to triples approximately equal to ϕ. The reward function $\Phi(\mathcal{G})$ is more in line with the sociological interpretation than the function given simply by the total number of triangles. The quantity ϕ should be interpreted as follows. If i is connected with both j and k, then j and k will be connected with probability ϕ. In the simulations to come we fix ϕ, choose a large value θ and use the Metropolis–Hastings algorithm to generate an outcome according to the measure P_θ^ϕ. Then the standard epidemic process is run on this graph in the usual way. For the rest of this section we assume that the neighbourhood size D is asymptotically Poisson distributed with parameter λ. We have not yet succeeded in finding expressions for the basic reproduction number of the model. However, in [69] a method is developed for finding a system of ordinary differential equations approximating the time development of the epidemic process. See also [48], [49], *et al.* [50] and [67]. We proceed to describe this method. (This pair approximation approach has also been used for lattice models by, for example, [62], [41], [71] and [70].)

5.1. Time dynamics

In order to find a way of approximating the time dynamics, we return to Equation (5). We should again close this system of equations at the level of pairs, but now the presence of triangles in the network adds a complication. If a given connected triple (i, j, k) is *not* part of a triangle, then as before the event that i is of type A is assumed to be independent of the event that k is of type C given that j is of type B. Thus

$$[abc] = \frac{[ab][bc]}{b} \qquad \text{for all } a, b, c. \tag{13}$$

On the other hand, if (i, j, k) does form a triangle then this relation has to be modified owing to the presence of the edge between i and k. We write instead

$$[abc] = \frac{[ab][bc]}{b}\Gamma_{ac} \qquad \text{for all } a, b, c, \tag{14}$$

where $\Gamma_{ac} = [ac]/(\lambda ac)$ should be interpreted as a measure of the correlation between individuals of type A and type C. For further details, see [69]. Finally, using (13) and (14) we insert

$$[abc] = \frac{[ab][bc]}{b}\left((1-\phi) + \phi\frac{[ac]}{\lambda ac}\right)$$

into the system (5) to obtain a closed system in the five variables x, y, $[xx]$, $[xy]$ and $[yy]$. The initial condition is given by $x(0) = 1-\bar{a}$, $y(0) = \bar{a}$, $[xx](0) = \lambda(1-\bar{a})^2$, $[xy](0) = \lambda\bar{a}(1-\bar{a})$ and $[yy](0) = \lambda\bar{a}^2$.

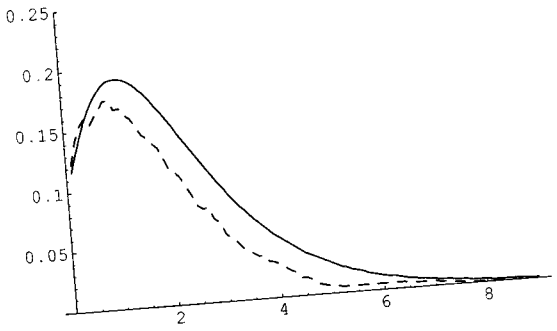

FIGURE 3: Deterministic vs. stochastic time dynamics.

TABLE 2: Deterministic final size τ and sample mean of the stochastic final size $1 - X(T)/n$.

ϕ	0.0		0.1		0.2	
β	det.	sim.	det.	sim.	det.	sim.
0.5	0.392	0.397	0.372	0.333	0.353	0.298
1.0	0.671	0.662	0.648	0.591	0.621	0.474
1.5	0.778	0.775	0.763	0.725	0.744	0.601
2.0	0.828	0.826	0.817	0.796	0.804	0.675
2.5	0.856	0.855	0.847	0.816	0.837	0.744

5.2. Simulations and numerical results

The numerical solution to this deterministic system will now be compared with simulations of the stochastic process. First, we have studied the time development of the relative size of infectious individuals. The average number of neighbours is given by $\lambda = 3$, the infection rate is $\beta = 1$, the proportion of initial infectives is $\bar{a} = 0.1$ and the interconnectedness is given by $\phi = 0.25$. The dashed line in Figure 3 gives a realization of the stochastic process $Y(t)/n$ with $n = 1000$ individuals and the solid line $y(t)$ is obtained from the solution to the deterministic equations.

Second, for different values of ϕ and β we have observed the deterministic approximation τ of the relative final size together with a sample mean of the corresponding random quantity, $1 - X(T)/n$. Here $\lambda = 3$, $\bar{a} = 0.1$, and in the simulations n has been set to 1000. The results are presented in Table 2, where for each pair (β, ϕ) the number on the left is the deterministic value and the value on the right is the estimate obtained from simulations. Thus the simulations show that both the relative number of infectious individuals and the final size proportion are overestimated by the deterministic system, and the discrepancy increases with ϕ. Nevertheless, the pair approximation approach certainly provides a qualitatively correct picture.

6. Overlapping subgraphs

When modelling the spread of disease in a human population, it is very important to take into account the formation of small social groups such as households, work groups and school classes. When considered separately, these structures decompose the population into mutually exclusive mixing groups, not allowing any disease spread worth mentioning. However, if the

structures are superimposed, a manifestly different situation arises. For instance, the epidemic may very well spread from one household to another via a transmission channel in the work group structure. The model of this section is motivated by such overlapping social networks; it also allows the superposition of a global structure representing the very irregular formation of circles of acquaintance.

Let us describe the model (cf. [7]). Let $\mathcal{G}^0, \mathcal{G}^1, \ldots, \mathcal{G}^r$ be subgraphs of the complete graph on n vertices. We assume that these different graphs have been generated by independent random mechanisms. Then form the union $\mathcal{G}$ in the obvious way. A given edge will belong to $\mathcal{G}$ if and only if it belongs to at least one of the graphs $\mathcal{G}^\rho$, $0 \le \rho \le r$. Denote by G^ρ (G) the adjacency matrix of $\mathcal{G}^\rho$ $(\mathcal{G})$. Finally, fix infection rates $\beta^0, \beta^1, \ldots, \beta^r$, one for each structure, and run an epidemic process on $\mathcal{G}$. A given infectious individual i makes close contacts with a given susceptible individual j at the points of a Poisson process with intensity $\sum_{\rho=0}^{r} \beta^\rho G_{ij}^\rho$, and the removal rate is equal to 1, as before. Since the graphs are assumed to be independent and sparse, when n is large typically at most one of the indicators G_{ij}^ρ, $0 \le \rho \le r$, will be non-zero.

We make further specific assumptions about the subgraphs $\mathcal{G}^\rho$. Assume that $\mathcal{G}^0$ is either a Bernoulli graph (Section 3) or a graph with prescribed degrees (Section 4). On the other hand, each of the graphs $\mathcal{G}^1, \ldots, \mathcal{G}^r$ is assumed to consist of disjoint complete subgraphs, corresponding to small mixing groups. With these simplifications some analytical results are possible, such as an expression for the basic reproduction number.

The special case where $\mathcal{G}^0$ is a Bernoulli graph, the average neighbourhood size λ tends to infinity and the infection rate β^0 tends to zero in such a way that $\lambda\beta^0$ is kept constant, and moreover $r = 1$, is referred to as the *household model*. For the household model, [14] is the main reference. For important contributions to the theory and practical applications, see e.g. [18], [19], [20] and [44]. A related model is treated in the early paper by Bartoszyński [16]. Work on outbreaks within households in the presence of community infection but without considering the dynamics of the latter can be found in [59], [1] and [17]. See also [9]. Finally, in an unpublished manuscript Altmann [5] discusses the case where $\mathcal{G}^0$ is absent and $r = 2$.

6.1. Basic reproduction number

By using branching approximations in a heuristic way, it is possible to derive the basic reproduction number R_0. We make the following modification of the generation process. When the disease enters a group, such as a household, it may take some generations before the local epidemic in the group has terminated. However, let us for simplicity pretend that the initially infectious individual always generates all secondary cases itself, so that the local epidemic is over within a single generation. The basic reproduction number is not affected by this assumption.

We say that an infectious individual is of 'type' ρ, $0 \le \rho \le r$, if it has become infected via an edge in $\mathcal{G}^\rho$. If the population is large, we may at the beginning approximate our generation process of infectives in terms of a multitype branching process $\mathbf{Y}(t) = (Y^0(t), \ldots, Y^\rho(t))$; $t = 0, 1, 2, \ldots$, where $Y^r(t)$ approximates the number of infectives of type ρ at generation t. The process has a special behaviour at time $t = 0$ as opposed to $t \ge 1$ (cf. Section 4), but this initial behaviour is irrelevant for the calculation of R_0. We only need to find the mean reproduction matrix Λ for the branching process $\mathbf{Y}(t)$, $t \ge 1$; then the largest eigenvalue of Λ will do as our basic reproduction number (cf. [42]).

First we need some definitions. Define

$$R(D) = \mathrm{E}(D) + \frac{\mathrm{var}(D) - \mathrm{E}(D)}{\mathrm{E}(D)},$$

where D is the (asymptotic) degree variable associated with $\mathcal{G}^0$. Put $R(D) = 0$ if $D \equiv 0$. Now consider one of the structures $\mathcal{G}^\rho$, $1 \leq \rho \leq r$. Let Π_k^ρ be the number of groups of size k in $\mathcal{G}^\rho$. Suppose that Π_k^ρ / n tends to π_k^ρ as n tends to infinity, where for mathematical convenience we assume that $\pi_k^\rho = 0$ for k large enough. Note that $\sum_{k=1}^\infty k\pi_k^\rho = 1$. Denote by m_k^ρ the average final size of a classical standard epidemic in a group of k individuals, where we start by infecting one single individual and where the infection rate is β^ρ. For the calculation of m_k^ρ, and indeed of the entire distribution of the final size, see [75] or [13]. The quantity

$$m^\rho = \sum_{k=1}^{\infty} k m_k^\rho \pi_k^\rho$$

gives the overall average of the final size. Thus, a randomly chosen individual will give rise to a local epidemic of average size m^ρ. Remember that the initial infective is included here, so that $m^\rho \geq 1$ always, with equality if and only if all the groups are of size 1.

An individual of type 0 has by definition been infected by a neighbour in $\mathcal{G}^0$ and will, according to the discussion of Section 4, infect on the average $p^0 R(D)$ new neighbours in $\mathcal{G}^0$, where $p^0 = \beta^0/(\beta^0 + 1)$. These individuals will have type 0 by definition. The individual will also infect on the average $m^\rho - 1$ new group members in $\mathcal{G}^\rho$ for each $\rho \geq 1$. Now consider instead a type ρ infective, $\rho \geq 1$. It has been infected by a group member in $\mathcal{G}^\rho$ and will thus not generate any new cases along this particular structure (the local epidemic is over within a single generation by assumption). On the other hand, for each $\rho' \neq \rho$, $\rho' \geq 1$, a mean number of $m^{\rho'} - 1$ group members in $\mathcal{G}^{\rho'}$ will become infected. Finally, on the average $p^0 E(D)$ neighbours in $\mathcal{G}^0$ will become infected by our type ρ infective, since all of these neighbours are susceptible! This yields the following mean reproduction matrix:

$$\Lambda = \begin{pmatrix} p^0 R(D) & m^1 - 1 & m^2 - 1 & \cdots \\ p^0 \mathrm{E}(D) & 0 & m^2 - 1 & \\ p^0 \mathrm{E}(D) & m^1 - 1 & 0 & \\ \vdots & & & \end{pmatrix}$$

The largest eigenvalue of the matrix Λ, i.e. the largest real root $\hat{\xi}$ of the equation $g(\xi) = 0$, where $g(\xi) = \det(\Lambda - \xi I)$, has in general no simple explicit expression. However, we only need conditions for $\hat{\xi}$ to be *below* 1 or *above* 1 in order to construct an appropriate expression for R_0 (see the definition given in Section 2) and such conditions may easily be found in some simple situations.

First, if $r = 0$ then $g(\xi) = p^0 R(D) - \xi$, so obviously $R_0 = \hat{\xi} = p^0 R(D)$ is the best choice (cf. Equation (10)).

As our next example we take $r = 1$. In this case

$$g(\xi) = \xi^2 - \xi p^0 R(D) - p^0 \mathrm{E}(D)(m_1 - 1).$$

If $D \equiv 0$, then $\hat{\xi} = 0$. Otherwise we note that g is a convex function with $g(0) \leq 0$ and $g'(0) < 0$, so that $\hat{\xi}$ is strictly positive. Since $\hat{\xi} < 1$ if and only if $g(1) > 0$ (we have passed the crossing point), a good candidate for the basic reproduction number is

$$R_0 = 1 - g(1) = p^0 \left(R(D) + (m^1 - 1)\mathrm{E}(D) \right). \tag{15}$$

Note that $\hat{\xi}$ itself is more complicated, containing a square root. When $\mathcal{G}^0$ has Poisson degrees (the Bernoulli graph model), we obtain $R_0 = m^1 \mathrm{E}(D)$ in accordance with the result given in [14].

Finally, assume that $r = 2$, $D \equiv 0$ (cf. [5]). As explained at the beginning of this section we may very well get a large outbreak even with a trivial structure $\mathcal{G}^0$, as long as there are (at least) two group structures interacting. We have in this case $g(\xi) = -\xi\left(\xi^2 - (m^1 - 1)(m^2 - 1)\right)$, so that a natural choice is

$$R_0 = (m^1 - 1)(m^2 - 1). \tag{16}$$

6.2. Final size equation

An equation for the asymptotic final size of the household model can also be derived. However, the notation becomes very cumbersome, and we refer to [14] for the details.

7. Dynamic graphs

In this final section we consider a simple social network that changes dynamically with time. When studying an epidemic over a large time period it is crucial to take into account the possibility of formation and dissolution of social relations. For instance, the spread of venereal diseases should preferably be modelled on a dynamic network of partnerships. The following simple construction due to Altmann [2] is amenable to mathematical analysis and at the same time reflects something of the behaviour expected of such a network.

As usual, we are given a set of n vertices. Suppose that two given non-adjacent vertices are joined by an edge at rate ρ^+/n, and adjacent vertices are disconnected at rate ρ^-. All these pair processes are assumed to be independent. In this way a simple continuous time graph process $\mathcal{G}(t)$, $t \geq 0$, is obtained. Assume that the process has reached equilibrium. Then the degree D_i of the ith vertex is binomially distributed with parameters $n-1$ and $\rho^+/(\rho^+ + n\rho^-)$, and thus converges in distribution to a Poisson variable with parameter $\lambda = \rho^+/\rho^-$ as n tends to infinity. The standard epidemic process is defined on this structure in the obvious way; at time $t \geq 0$ we simply follow the transition rules induced by the graph $\mathcal{G}(t)$. For possible generalizations of this model, see [2]. Dynamic partnership models have also been discussed in [30], [31]), [34] and [64]. See also [24].

7.1. Basic reproduction number

The basic reproduction number is easy to calculate. Consider a single infectious individual in a totally susceptible population. The probability that this individual and a given susceptible individual are partners at the beginning of the infectious period, or else become partners during this time, is given by

$$\frac{\rho^+/n}{\rho^+/n + \rho^-} \cdot 1 + \left(1 - \frac{\rho^+/n}{\rho^+/n + \rho^-}\right) \cdot \frac{\rho^+/n}{1 + \rho^+/n}.$$

Given this event, infection will occur with probability $\beta/(1+\beta+\rho^-)$. Note that the probability of more than one partnership episode between these two individuals during the infectious period may be neglected. Thus the expected number of new cases is given by $n-1$ times the product of these two probabilities. As n grows, this quantity tends to

$$R_0 = \frac{\rho^+\beta(1+\rho^-)}{\rho^-(1+\beta+\rho^-)} = \frac{\lambda\beta(1+\rho^-)}{1+\beta+\rho^-}. \tag{17}$$

Note that if the rates ρ^+ and ρ^- tend to zero while their ratio λ is kept constant, the basic reproduction number tends to the value given in the Bernoulli graph case, see Equation (4). This is to be expected since the network is then made increasingly static. Also, since the expression in (17) *decreases* to the limit in (4) as the dissolution rate tends to zero, we learn that less frequent change of partners will diminish the spread of disease.

7.2. Time dynamics

The asymptotic time dynamics of the epidemic process is derived rigorously in [4]. Assume that the proportion of initial infectives tends to $\bar{a}$, $0 < \bar{a} < 1$. The resulting differential system is as follows:

$$\begin{aligned} \frac{dx}{dt} &= -\beta x \hat{y}, \\ \frac{dy}{dt} &= \beta x \hat{y} - y, \\ \frac{d\hat{y}}{dt} &= \lambda\beta x \hat{y} - (\beta + 1)\hat{y} + \rho^+ y - \rho^- \hat{y}, \end{aligned} \tag{18}$$

with initial condition $x(0) = 1 - \bar{a}$, $y(0) = \bar{a}$ and $\hat{y}(0) = \lambda\bar{a}$. Using the results of Section 3, it is easy to understand why this system is correct. If we put $\rho^+ = \rho^- = 0$, we obtain the system (7). With dynamic partnerships the following happens. Each given disconnected pair initiate new relations at rate ρ^+/n, hence we see that $[xy]$ increases at rate $\rho^+ xy$. Also, dissolution of such edges occurs at rate $\rho^-[xy]$, leading to Equation (18).

7.3. Final size equation

Following [4], an equation for the asymptotic final size of the epidemic, or the asymptotic escaping proportion, is obtained next. Consider the system (18). Integrating the first equation yields

$$\sigma = (1 - \bar{a}) \exp\left\{-\beta \int_0^\infty \hat{y}(t)\, dt\right\}. \tag{19}$$

Then, by summing the first two equations and integrating we get

$$\sigma = 1 - \int_0^\infty y(t)\, dt. \tag{20}$$

Finally, by writing the third equation as

$$\frac{d\hat{y}}{dt} = \lambda\frac{dy}{dt} + \lambda(1 + \rho^-)y - (1 + \beta + \rho^-)\hat{y}$$

and integrating we see that

$$0 = \lambda(1 + \rho^-)\int_0^\infty y(t)\, dt - (1 + \beta + \rho^-)\int_0^\infty \hat{y}(t)\, dt. \tag{21}$$

Combining (19), (20) and (21), we finally find

$$\sigma = (1 - \bar{a}) \exp\left\{-\frac{\lambda\beta(1 + \rho^-)}{1 + \beta + \rho^-}(1 - \sigma)\right\} = (1 - \bar{a})e^{-R_0(1-\sigma)},$$

i.e. the classical final size equation. Not only does more frequent change of partners increase the basic reproduction number, but the final size is similarly increased.

8. Discussion

The prime message of our presentation of stochastic epidemic models on graphs and networks has been to highlight the importance of allowing repeated contacts between individuals in models for the spread of disease. The concept of heterogeneous mixing as introduced in this text has a fundamental influence on the basic modelling assumptions, whereas other effects caused by multitype populations, general infectious periods, births and deaths, immigration and emigration, age-dependence, partial or temporary immunity, change of behaviour, and so on, may presumably be incorporated at later stages.

It is clear that the mathematical analysis is made considerably more difficult when the networks contain short loops (an exception being the formation of small disjoint groups with complete mixing as presented in Section 6); nevertheless it is crucial to find ways of attacking these more realistic model structures. We are currently investigating possible ways of deriving the basic reproduction number for a network with Poisson degrees and a certain number of triangles; that is, the network considered in Section 5.

We have so far presented various ways to model *social relations* in a population, thus ignoring the possibility of geographical spread of the disease. For the standard epidemic model on the two-dimensional lattice, the spatial element completely determines the progress of the epidemic. This model has been studied in [65], [55], [57] and [27]. For an excellent introduction to the general theory of interacting particle systems, see Durrett [35]. The book by Liggett [58] also provides a thorough treatment of the subject, and [36] gives a survey oriented towards ecological applications.

In order to approach reality, social space and geographical space should be considered simultaneously, but to find natural models remains an open problem. Interesting ideas, applicable mainly in a meta-population situation, are given by Cliff [26], who deforms geographical space by a non-linear mapping in such a way that the usual distance metric is transformed into an epidemiological metric (e.g. epidemic time-lags measured in some time unit). The 'great circle model' in [14], although highly idealized, also is of interest in this respect. Hopefully models for epidemic spread on social networks with increasingly realistic properties will continue to be investigated in future.

References

[1] ADDY, C. L., LONGINI I. M. AND HABER M. (1991). A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* **47**, 961–974.

[2] ALTMANN, M. (1995). Susceptible-Infected-Removed epidemic models with dynamic partnerships. *J. Math. Biol.* **33**, 661–675.

[3] ALTMANN, M. (1996). Network measures for epidemiology. In *Models for Infectious Human Diseases: Their Structure and Relation to Data*, eds. V. Isham and G. Medley. Publications of the Newton Institute, Cambridge University Press.

[4] ALTMANN, M. (1998). The deterministic limit of infectious disease models with dynamic partners. *Math. Biosci.* **150**, 153–175.

[5] ALTMANN, M. Epidemic models with overlapping mixing groups. Unpublished manuscript.

[6] ANDERSON, R. M. AND MAY, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford University Press.

[7] ANDERSSON, H. (1997). Epidemics in a population with social structures. *Math. Biosci.* **140**, 79–84.

[8] ANDERSSON, H. (1998). Limit theorems for a random graph epidemic model. *Ann. Appl. Prob.* **8**, 1331–1349.

[9] ANDERSSON, H. AND BRITTON, T. (1998). Heterogeneity in epidemic models and its effect on the spread of infection. *J. Appl. Prob.* **35**, 651–661.

[10] VON BAHR, B. AND MARTIN-LÖF, A. (1980). Threshold limit theorems for some epidemic processes. *Adv. Appl. Prob.* **12**, 319–349.

[11] BAILEY, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika* **40**, 177–185.

[12] BAILEY, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases*. Griffin, London.
[13] BALL, F. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Prob.* **18**, 289–310.
[14] BALL, F., MOLLISON, D. AND SCALIA-TOMBA, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–89.
[15] BARBOUR, A. D. AND MOLLISON, D. (1990). Epidemics and random graphs. In *Stochastic processes in epidemic theory*, eds. J.P. Gabriel, C. Lefèvre and P. Picard (Lecture Notes in Biomath. **86**). Springer, Berlin, pp. 86–89.
[16] BARTOSZYŃSKI, R. (1972). On a certain model of an epidemic. *Applicationes Mathematicae* **13**, 139–151.
[17] BECKER, N. G. (1989). *Analysis of Infectious Disease Data*. Chapman and Hall, London.
[18] BECKER, N. G. AND DIETZ, K. (1995). The effect of the household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127**, 207–219.
[19] BECKER, N. G. AND HALL, R. (1996). Immunization levels for preventing epidemics in a community of households made up of individuals of different types. *Math. Biosci.* **132**, 205–216.
[20] BECKER, N. G. AND STARCZAK, D. N. (1997). Optimal vaccination strategies for a community of households. *Math. Biosci.* **139**, 117–132.
[21] BENDER, E. A. AND CANFIELD, E. R. (1978). The asymptotic number of labelled graphs with given degree sequences. *J. Combinat. Theory* (A) **24**, 296–307.
[22] BERG, S. (1981). On snowball sampling, random mappings and related problems. *J. Appl. Prob.* **18**, 283–290.
[23] BERG, S. (1983). Random contact processes, snowball sampling and factorial series distributions. *J. Appl. Prob.* **20**, 31–46.
[24] BLANCHARD, P., BOLZ, G. F. AND KRÜGER, T. (1990). Modelling AIDS-epidemics or any venereal disease on random graphs. In *Stochastic Processes in Epidemic Theory*, eds. J. P. Gabriel, C. Lefèvre and P. Picard (Lecture Notes in Biomath. **86**). Springer, Berlin, pp. 104–117.
[25] BOLLOBÁS, B (1985). *Random Graphs*. Academic Press, New York.
[26] CLIFF, A. (1995). Incorporating spatial components into models of epidemic spread. In *Epidemic models: Their Structure and Relation to Data*, ed. D. Mollison. Publications of the Newton Institute, Cambridge University Press.
[27] COX, J. T. AND DURRETT, R. (1988). Limit theorems for the spread of epidemics and forest fires. *Stoch. Proc. Appl.* **30**, 171–191.
[28] DIEKMANN, O., DE JONG, M. C. M. AND METZ, J. A. J. (1998). A deterministic epidemic model taking account of repeated contacts between the same individuals. *J. Appl. Prob.* **35**, 448–462.
[29] DIEKMANN, O. HEESTERBEEK, J. A. P. AND METZ, J. A. J. (1990). On the definition and the computation of the basic reproduction number R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.
[30] DIETZ, K. (1988). On the transmission dynamics of HIV. *Math. Biosci.* **90**, 397–414.
[31] DIETZ, K. (1988). The dynamics of spread of HIV infection in the heterosexual population. In *Statistical analysis and mathematical modelling of AIDS*, eds. J. C. Jager and E. J. Ruitenberg. Oxford University Press.
[32] DIETZ, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statist. Meth. Med. Res.* **2**, 23–41.
[33] DIETZ, K. (1995). Some problems in the theory of infectious disease transmission and control. In *Epidemic models: Their Structure and Relation to Data*, ed. D. Mollison. Publications of the Newton Institute, Cambridge University Press.
[34] DIETZ, K. AND HADELER, K. P. (1988). Epidemiological models for sexually transmitted diseases. *J. Math. Biol.* **26**, 1–25.
[35] DURRETT, R. (1995). Ten lectures on particle systems. *Lecture Notes in Math.* **1608**, 97–201.
[36] DURRETT, R. AND LEVIN, S. A. (1994). Stochastic spatial models: A user's guide to ecological applications. *Phil. Trans. R. Soc. London B* **343**, 329–350.
[37] ERDÖS, P. AND RÉNYI, A. (1959). On random graphs I. *Publ. Math. Debrecen* **6**, 290–297.
[38] ETHIER, S. N. AND KURTZ, T. G. (1986). *Markov processes, characterization and convergence*. Wiley, New York.
[39] FRANK, O. AND STRAUSS, D. (1986). Markov graphs. *J. Amer. Stat. Assoc.* **81**, 832–842.
[40] GERTSBAKH, I. B. (1977). Epidemic process on a random graph: Some preliminary results. *J. Appl. Prob.* **14**, 427–438.
[41] HARADA, Y. AND ISAWA, Y. (1994). Lattice population dynamics for plants with dispersing seeds and vegetative propagation. *Res. Popul. Ecol.* **36**, 237–249.
[42] HARRIS, T. E. (1963). *The Theory of Branching Processes*. Dover, New York.
[43] HEESTERBEEK, J. A. P. (1992). R_0, Ph.D. thesis, Centrum voor Wiskunde en Informatica, Amsterdam.
[44] ISLAM, M. N., O'SHAUGHNESSY, C. D. AND SMITH, B. (1996). A random graph model for the final-size distribution of household infections. *Statistics in Medicine* **15**, 837–843.

[45] JAWORSKI, J. *Alg.* **5**, 73–94.
[46] JAWORSKI, J.
[47] JONASSON, J.
[48] KEELING, M. University, Cov
[49] KEELING, M. ecologies. In *systems'*, ed. P.
[50] KEELING, M. *Proc. R. Soc. L*
[51] KESTEN, H. (
[52] KOLCHIN, V. I
[53] KURTZ, T. G. differential pro
[54] KURTZ, T. G.
[55] KUULASMAA, **19**, 745–758.
[56] KUULASMAA, **1212**, 195–199
[57] KUULASMAA, *J. Appl. Prob.*
[58] LIGGETT, T. M
[59] LONGINI, I. M distributions of
[60] LUDWIG, D. (
[61] MARTIN-LÖF, limit theorems
[62] MATSUDA, H. Lotka–Volterra
[63] MCKAY, B. D **19A**, 15–25.
[64] MODE, C. J. Semi-Markov
[65] MOLLISON, D 283–326.
[66] MOLLOY, M. *Struct. Alg.* **6**,
[67] MORRIS, A. University, C
[68] PICARD, P. A Reed–Frost e
[69] RAND, D. *Ecology* 2, e
[70] SATŌ, K. AI *Soc. Japan* **6**
[71] SATŌ, K., M populations
[72] STEPANOV 55–67.
[73] STRAUSS,
[74] WATSON, *Prob.* **17**, 3
[75] P. WHITT 122.

[45] JAWORSKI, J. AND MUTAFCHIEV, L. (1994). The largest component in a random mapping. *Random Struct. Alg.* **5**, 73–94.
[46] JAWORSKI, J. AND SMIT, I. H. (1987). On a random digraph. *Ann. Discrete Math.* **33**, 111–127.
[47] JONASSON, J. (1999). The random triangle model. *J. Appl. Prob* **36**, 852–867
[48] KEELING, M. J. (1995). The ecology and evolution of spatial host–parasite systems. Ph.D. thesis, Warwick University, Coventry, UK.
[49] KEELING, M. J. AND RAND, D.A. (1996). Spatial correlations and local fluctuations in host–parasite ecologies. In *Proceedings of the Isaac Newton Institute Programme 'From finite to infinite dimensional systems'*, ed. P. Glendinning. Kluwer, Dordrecht.
[50] KEELING, M. J., RAND, D. A. AND MORRIS, A. J. (1997). Correlation models for childhood epidemics. *Proc. R. Soc. London B* **264**, 1149–1156.
[51] KESTEN, H. (1982). *Percolation Theory for Mathematicians*. Birkhauser, Boston.
[52] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, New York.
[53] KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Prob.* **8**, 344–356.
[54] KURTZ, T. G. (1981). *Approximation of Population Processes*. SIAM, Philadelphia, PA.
[55] KUULASMAA, K. (1982). The spatial general epidemic and locally dependent random graphs. *J. Appl. Prob.* **19**, 745–758.
[56] KUULASMAA, K. (1986). On the reproduction rate of the spatial general epidemic. *Lecture Notes in Math.* **1212**, 195–199.
[57] KUULASMAA, K. AND ZACHARY, S. (1984). On spatial general epidemics and bond percolation processes. *J. Appl. Prob.* **21**, 911–914.
[58] LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.
[59] LONGINI, I. M. AND KOOPMAN, J. S. (1982). Household and community transmission parameters from final distributions of infections in households. *Biometrics* **38**, 115–126.
[60] LUDWIG, D. (1974). *Stochastic Population Theories* (Lecture notes in Biomath. **3**). Springer, Berlin.
[61] MARTIN-LÖF, A. (1986). Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Prob.* **23**, 265–282.
[62] MATSUDA, H., OGITA, N., SASAKI, A. AND SATŌ, K. (1992). Statistical mechanics of population–The lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88**, 1035–1049.
[63] MCKAY, B. D. (1985). Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combinatorica* **19A**, 15–25.
[64] MODE, C. J. AND DIETZ, K. (1994). On some formulas in a partnership model from the perspective of a Semi-Markov process. *J. Math. Biol.* **32**, 161–169.
[65] MOLLISON, D. (1977). Spatial contact models for ecological and epidemic spread. *J. Roy. Statist. Soc. B* **39**, 283–326.
[66] MOLLOY, M. AND REED, B. (1995). A critical point for random graphs with a given degree sequence. *Random Struct. Alg.* **6**, 161–179.
[67] MORRIS, A. J. (1997). Representing spatial interactions in simple ecological models. Ph.D. thesis, Warwick University, Coventry, UK.
[68] PICARD, P. AND LEFÈVRE, C. (1990). A unified analysis of the final size and severity distribution in collective Reed–Frost epidemic processes. *Adv. Appl. Prob.* **22**, 269–294.
[69] RAND, D. A. (1997). Correlation equations and pair approximations for spatial ecologies. In *Theoretical Ecology* 2, ed. J. McGlade. Blackwell Scientific Publishing, Oxford.
[70] SATŌ, K. AND KONNO N. (1995). Successional dynamical models on the 2-dimensional lattice space. *J. Phys. Soc. Japan* **64**, 1866–1869.
[71] SATŌ, K., MATSUDA, H. AND SASAKI, A. (1994). Pathogen invasion and host extinction in lattice structured populations. *J. Math. Biol.* **32**, 251–268.
[72] STEPANOV, V. E. (1970). On the probability of connectedness of a random graph. *Theory Prob. Appl.* **15**, 55–67.
[73] STRAUSS, D. (1986). On a general class of models for interaction. *SIAM Review* **28**, 513–527.
[74] WATSON, R. K. (1980). A useful random time-scale transformation for the standard epidemic model. *J. Appl. Prob.* **17**, 324–332.
[75] P. WHITTLE, P. (1955). The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika* **42**, 116–122.