Chapter 4
An Introduction to Networks in Epidemic Modeling

Fred Brauer

Abstract We use a stochastic branching process to describe the beginning of a disease outbreak. Unlike compartmental models, if the basic reproduction number is greater than one there may be a minor outbreak or a major epidemic with a probability depending on the nature of the contact network. We use a network approach to determine the distribution of outbreak and epidemic sizes.

4.1 Introduction

The Kermack–McKendrick compartmental epidemic model assumes that the sizes of the compartments are large enough that the mixing of members is homogeneous, or at least that there is homogeneous mixing in each subgroup if the population is stratified by activity levels. However, at the beginning of a disease outbreak, there is a very small number of infective individuals and the transmission of infection is a stochastic event depending on the pattern of contacts between members of the population; a description should take this pattern into account.

It has often been observed in epidemics that there is a small number of "superspreaders" who transmit infection to many other members of the population, while most infectives do not transmit infections at all or transmit infections to very few others [17]. This suggests that homogeneous mixing at the beginning of an epidemic may not be a good assumption. The SARS epidemic of 2002–2003 spread much more slowly than would have been expected on the basis of the data on disease spread at the start of the epidemic. Early

Department of Mathematics, University of British Columbia, 1984, Mathematics Road, Vancouver BC, Canada V6T 1Z2 brauer@math.ubc.ca

in the SARS epidemic of 2002–2003 it was estimated that $\mathcal{R}_0$ had a value between 2.2 and 3.6. At the beginning of an epidemic, the exponential rate of growth of the number of infectives is approximately $(\mathcal{R}_0 - 1)/\alpha$, where $1/\alpha$ is the generation time of the epidemic, estimated to be approximately 10 days for SARS . This would have predicted at least $30,000$ cases of SARS in China during the first four months of the epidemic. In fact, there were fewer than 800 cases reported in this time. An explanation for this discrepancy is that the estimates were based on transmission data in hospitals and crowded apartment complexes. It was observed that there was intense activity in some locations and very little in others. This suggests that the actual reproduction number (averaged over the whole population) was much lower, perhaps in the range 1.2–1.6, and that heterogeneous mixing was a very important aspect of the epidemic.

4.2 The Probability of a Disease Outbreak

Our approach will be to give a stochastic branching process description of the beginning of a disease outbreak to be applied so long as the number of infectives remains small, distinguishing a (minor) disease outbreak confined to this stage from a (major) epidemic which occurs if the number of infectives begins to grow at an exponential rate . Once an epidemic has started we may switch to a deterministic compartmental model, arguing that in a major epidemic contacts would tend to be more homogeneously distributed. However, if we continue to follow the network model we would obtain a somewhat different estimate of the final size of the epidemic. Simulations suggest that the assumption of homogeneous mixing in a compartmental model may lead to a higher estimate of the final size of the epidemic than the prediction of the network model.

We describe the network of contacts between individuals by a graph with members of the population represented by vertices and with contacts between individuals represented by edges. The study of graphs originated with the abstract theory of Erdös and Rényi of the 1950s and 1960s [3–5], and has become important more recently in many areas, including social contacts and computer networks, as well as the spread of communicable diseases. We will think of networks as bi-directional, with disease transmission possible in either direction along an edge.

An edge is a contact between vertices that can transmit infection. The number of edges of a graph at a vertex is called the *degree* of the vertex. The degree distribution of a graph is $\{p_k\}$, where p_k is the fraction of vertices having degree k. The degree distribution is fundamental in the description of the spread of disease. Initially, we assume that all contacts between an infective and a susceptible transmit infection, but we will relax this assumption in Sect. 4.3.

We think of a small number of infectives in a population of susceptibles large enough that in the initial stage we may neglect the decrease in the size of the susceptible population. Our development begins along the lines of that of [7] and then develops along the lines of [6, 14, 16]. We assume that the infectives make contacts independently of one another and let p_k denote the probability that the number of contacts by a randomly chosen individual is exactly k, with $\sum_{k=0}^{\infty} p_k = 1$. In other words, $\{p_k\}$ is the degree distribution of the vertices of the graph corresponding to the population network.

For convenience, we define the *generating function*

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Since $\sum_{k=0}^{\infty} p_k = 1$, this power series converges for $0 \leq z \leq 1$, and may be differentiated term by term. Thus

$$p_k = \frac{G_0^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \cdots .$$

It is easy to verify that the generating function has the properties

$$G_0(0) = p_0, \quad G_0(1) = 1, \quad G_0'(z) > 0, \quad G_0''(z) > 0.$$

The mean degree, which we denote by $< k >$, is

$$< k >= \sum_{k=1}^{\infty} k p_k = G_0'(1).$$

More generally, we define the moments

$$< k^j >= \sum_{k=1}^{\infty} k^j p_k, \quad j = 1, 2, \cdots \infty.$$

When a disease is introduced into a network, we think of it as starting at a vertex (patient zero) who transmits infection to every individual to whom this individual is connected, that is, along every edge of the graph from the vertex corresponding to this individual. We assume that this initial vertex has been infected by a contact outside the population (component of the network) being studied. For transmission of disease after this initial contact we need to use the *excess degree* of a vertex. If we follow an edge to a vertex, the excess degree of this vertex is one less than the degree. We use the excess degree because infection can not be transmitted back along the edge whence it came. The probability of reaching a vertex of degree k, or excess degree $(k-1)$, by following a random edge is proportional to k, and thus the probability that a vertex at the end of a random edge has excess degree $(k-1)$ is a constant multiple of kp_k with the constant chosen to make the sum over k

of the probabilities equal to 1. Then the probability that a vertex has excess degree $(k-1)$ is

$$q_{k-1} = \frac{kp_k}{< k >}.$$

This leads to a generating function $G_1(z)$ for the excess degree

$$G_1(z) = \sum_{k=1}^{\infty} q_{k-1} z^{k-1} = \sum_{k=1}^{\infty} \frac{kp_k}{< k >} z^{k-1} = \frac{1}{< k >} G_0'(z),$$

and the mean excess degree, which we denote by $< k_e >$, is

$$\begin{aligned} < k_e > &= \frac{1}{< k >} \sum_{k=1}^{\infty} k(k-1)p_k \\ &= \frac{1}{< k >} \sum_{k=1}^{\infty} k^2 p_k - \frac{1}{< k >} \sum_{k=1}^{\infty} kp_k \\ &= \frac{< k^2 >}{< k >} - 1 = G_1'(1). \end{aligned}$$

We let $\mathcal{R}_0 = G_1'(1)$, the mean excess degree. This is the mean number of secondary cases infected by patient zero and is the basic reproduction number as usually defined; the threshold for an epidemic is determined by $\mathcal{R}_0$.

Our next goal is to calculate the probability that the infection will die out and will not develop into a major epidemic. We begin by assuming that patient zero is a vertex of degree k of the network. Suppose patient zero transmits infection to a vertex of degree j. We let z_n denote the probability that this infection dies out within the next n generations. For the infection to die out in n generations each of these j secondary infections must die out in $(n-1)$ generations. The probability of this is z_{n-1} for each secondary infection, and the probability that all secondary infections will die out in $(n-1)$ generations is z_{n-1}^j. Now z_n is the sum over j of these probabilities, weighted by the probability q_j of j secondary infections. Thus

$$z_n = \sum_{j=0}^{\infty} q_j z_{n-1}^j = G_1(z_{n-1}).$$

Since $G_1(z)$ is an increasing function, the sequence z_n is an increasing sequence and has a limit z_∞, which is the probability that this infection will die out eventually. Then z_∞ is the limit as $n \to \infty$ of the solution of the difference equation

$$z_n = G_1(z_{n-1}), \quad z_0 = 0.$$

Thus z_∞ must be an equilibrium of this difference equation, that is, a solution of $z = G_1(z)$. Let w be the smallest positive solution of $z = G_1(z)$. Then, because $G_1(z)$ is an increasing function of z, $z \le G_1(z) \le G_1(w) = w$ for

$0 \leq z \leq w$. Since $z_0 = 0 < w$ and $z_{n-1} \leq w$ implies

$$z_n = G_1(z_{n-1}) \leq G_1(w) = w,$$

it follows by induction that

$$z_n \leq w,\ n = 0, 1, \cdots \infty.$$

From this we deduce that

$$z_\infty = w.$$

The equation $G_1(z) = z$ has a root $z = 1$ since $G_1(1) = 1$. Because the function $G_1(z) - z$ has a positive second derivative, its derivative $G_1'(z) - 1$ is increasing and can have at most one zero. This implies that the equation $G_1(z) = z$ has at most two roots in $0 \leq z \leq 1$. If $\mathcal{R}_0 < 1$ the function $G_1(z) - z$ has a negative first derivative

$$G_1'(z) - 1 \leq G_1'(1) - 1 = \mathcal{R}_1 - 1 < 0$$

and the equation $G_1(z) = z$ has only one root, namely $z = 1$. On the other hand, if $\mathcal{R}_0 > 1$ the function $G_1(z) - z$ is positive for $z = 0$ and negative near $z = 1$ since it is zero at $z = 1$ and its derivative is positive for $z < 1$ and z near 1. Thus in this case the equation $G_1(z) = z$ has a second root $z_\infty < 1$.

The probability that the disease outbreak will die out eventually is the sum over k of the probabilities that the initial infection in a vertex of degree k will die out, weighted by the degree distribution $\{p_k\}$ of the original infection, and this is

$$\sum_{k=0}^{\infty} p_k z_\infty^k = G_0(z_\infty).$$

To summarize this analysis, we see that if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1. On the other hand, if $\mathcal{R}_0 > 1$ there is a solution $z_\infty < 1$ of

$$G_1(z) = z$$

and there is a probability $1 - G_0(z_\infty) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $G_0(z_\infty)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic. This distinction between a minor outbreak and a major epidemic, and the result that if $\mathcal{R}_0 > 1$ there may be only a minor outbreak and not a major epidemic are aspects of stochastic models not reflected in deterministic models.

4.3 Transmissibility

Contacts do not necessarily transmit infection. For each contact between individuals of whom one has been infected and the other is susceptible there is a probability that infection will actually be transmitted. This probability depends on such factors as the closeness of the contact, the infectivity of the member who has been infected, and the susceptibility of the susceptible member. We assume that there is a mean probability T, called the *transmissibility*, of transmission of infection. The transmissibility depends on the rate of contacts, the probability that a contact will transmit infection, the duration time of the infection, and the susceptibility. The development in Sect. 4.2 assumed that all contacts transmit infection, that is, that $T = 1$.

In this section, we will continue to assume that there is a network describing the contacts between members of the population whose degree distribution is given by the generating function $G_0(z)$, but we will assume in addition that there is a mean transmissibility T.

When disease begins in a network, it spreads to some of the vertices of the network. Edges that are infected during a disease outbreak are called *occupied*, and the size of the disease outbreak is the cluster of vertices connected to the initial vertex by a continuous chain of occupied edges.

The probability that exactly m infections are transmitted by an infective vertex of degree k is

$$\binom{k}{m} T^m (1-T)^{k-m}.$$

We define $\mathit{\Gamma}_0(z,T)$ be the generating function for the distribution of the number of occupied edges attached to a randomly chosen vertex, which is the same as the distribution of the infections transmitted by a randomly chosen individual for any (fixed) transmissibility T. Then

$$\begin{aligned}\mathit{\Gamma}_0(z,T) &= \sum_{m=0}^{\infty}\left[\sum_{k=m}^{\infty} p_k \binom{k}{m} T^m (1-T)^{(k-m)}\right] z^m \\ &= \sum_{k=0}^{\infty} p_k \left[\sum_{m=0}^{k} \binom{k}{m} (zT)^m (1-T)^{(k-m)}\right] \\ &= \sum_{k=0}^{\infty} p_k [zT + (1-T)]^k = G_0(1 + (z-1)T).\end{aligned} \tag{4.1}$$

In this calculation we have used the binomial theorem to see that

$$\sum_{m=0}^{k} \binom{k}{m} (zT)^m (1-T)^{(k-m)} = [zT + (1-T)]^k.$$

Note that

$$\Gamma_0(0,T) = G_0(1-T), \quad \Gamma_0(1,T) = G_0(1) = 1, \quad \Gamma_0'(z,T) = TG_0'(1+(z-1)T).$$

For secondary infections we need the generating function $\Gamma_1(z,T)$ for the distribution of occupied edges leaving a vertex reached by following a randomly chosen edge. This is obtained from the excess degree distribution in the same way,

$$\Gamma_1(z,T) = G_1(1 + (z-1)T)$$

and

$$\Gamma_1(0,T) = G_1(1-T), \quad \Gamma_1(1,T) = G_1(1) = 1, \quad \Gamma_1'(z,T) = TG_1'(1+(z-1)T).$$

The basic reproduction number is now

$$\mathcal{R}_0 = \Gamma_1'(1,T) = TG_1'(1).$$

The calculation of the probability that the infection will die out and will not develop into a major epidemic follows the same lines as the argument in Sect. 4.2 for $T = 1$. The result is that if $\mathcal{R}_0 = TG_1'(1) < 1$ the probability that the infection will die out is 1. If $\mathcal{R}_0 > 1$ there is a solution $z_\infty(T) < 1$ of

$$\Gamma_1(z,T) = z,$$

and a probability $1-\Gamma_0(z_\infty(T),T) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $\Gamma_1(z_\infty(T),T)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

Another interpretation of the basic reproduction number is that there is a *critical transmissibility* T_c defined by

$$T_c G_1'(1) = 1.$$

In other words, the critical transmissibility is the transmissibility that makes the basic reproduction number equal to 1. If the mean transmissibility can be decreased below the critical transmissibility, then an epidemic can be prevented.

The measures used to try to control an epidemic may include contact interventions, that is, measures affecting the network such as avoidance of public gatherings and rearrangement of the patterns of interaction between caregivers and patients in a hospital, and transmission interventions such as careful hand washing or face masks to decrease the probability that a contact will lead to disease transmission.

4.4 The Distribution of Disease Outbreak and Epidemic Sizes

We define $H_0(z,T)$ to be the generating function for the distribution of outbreak sizes corresponding to a randomly chosen vertex. In a corresponding way, we define $H_1(z,T)$ to be the generating function for the sizes of the clusters of connected vertices reached by following a randomly chosen edge.

For the generating function $H_1(z,T)$, it is easy to verify that $[H_1(z,T)]^2$ represents the distribution function for the sum of the infected cluster sizes for two vertices, and similarly for higher powers. If we begin on a randomly chosen edge, the probability that the vertex at the end of this edge has degree k is q_k, and each of the k vertices connected to it has a distribution of infected cluster sizes given by $H_1(z,T)$. Then

$$\begin{aligned} H_1(z,T) &= z\sum_{m=0}^{\infty}\left[\sum_{k=m}^{\infty} q_k\binom{k}{m}T^m(1-T)^{(k-m)}\right](H_1(z,T))^m \\ &= z\sum_{k=0}^{\infty} q_k\left[\sum_{m=0}^{k}\binom{k}{m}(TH_1(z,T))^m(1-T)^{(k-m)}\right] \\ &= z\sum_{k=0}^{\infty} q_k[TH_1(z,T)+(1-T)]^k = zG_1(1+(H_1(z,T)-1)T). \end{aligned}$$

Thus

$$H_1(z,T) = z\Gamma_1(H_1(z,T),T). \tag{4.2}$$

Similarly, the size of the cluster reachable from a randomly chosen vertex is distributed according to

$$H_0(z,T) = z\Gamma_0(H_1(z,T),T). \tag{4.3}$$

The mean size of the disease outbreak is $H_0'(1,T)$. We calculate this by implicit differentiation of (4.2) after using implicit differentiation of (4.3) to calculate $H_1'(z,t)$.

Implicit differentiation of (4.2) gives

$$\begin{aligned} H_1'(z,T) &= \Gamma_1(H_1(z,T),T) + z\Gamma_1'(H_1(z,T),T)H_1'(z,T) \\ &= \frac{\Gamma_1(H_1(z,T),T)}{1-z\Gamma_1'(H_1(z,T),T)} \\ H_1'(1,T) &= \frac{\Gamma_1(H_1(1,T),T)}{1-\Gamma_1'(H_1(1,T),T)}. \end{aligned} \tag{4.4}$$

Then implicit differentiation of (4.3) using (4.4) gives

$$H_0'(z,T) = \Gamma_0(H_1(z,T),T) + z\Gamma_0'(H_1(z,T),T)H_1'(z,T) \tag{4.5}$$
$$= \Gamma_0(H_1(z,T),T) + z\Gamma_0'(H_1(z,T),T)\frac{\Gamma_1(H_1(z,T),T)}{1 - z\Gamma_1'(H_1(z,T),T)}.$$

Because

$$H_1(1,T) = 1, \quad \Gamma_1(H_1(1,T),T) = \Gamma_1(1,T) = 1,$$

this reduces to

$$H_0'(1,T) = 1 + \frac{\Gamma_0'(1,T)}{1 - \Gamma_1'(1,T)} = 1 + \frac{TG_0'(1)}{1 - TG_1'(1)} = 1 + \frac{TG_0'(1)}{1 - \mathcal{R}_0}.$$

This expression for the mean outbreak size is valid if $\mathcal{R}_0 = TG_1'(1) < 1$.

There is a phase transition at $\mathcal{R}_0 = 1$. A "giant" component of the graph appears, and there is a major epidemic. If $\mathcal{R}_0 \geq 1$, we exclude the "giant" component of the graph from the definition of $H_1(z,T)$ and then $H_1(1,T) < 1$. Because of (4.2) we must have

$$H_1(1,T) = \Gamma_1(H_1(1,T))$$

and therefore $H_1(1,T)$ must be the second root $z_\infty(T)$ of

$$\Gamma_1(z,T) = z$$

as found in Sect. 4.3. In this case, $\Gamma_0(z_\infty(T))$ is the probability that there will be only a small disease outbreak and $1 - \Gamma_0(z_\infty(T))$ is the probability that there will be an epidemic.

If $\mathcal{R}_0 < 1$, $H_1(1,T) = 1, z_\infty(T) = 1$, and the probability of an epidemic is 0. If there is an epidemic, we define $S(T)$ to be the fraction of the graph affected by the infection, the epidemic size. Above the epidemic threshold,

$$H_0(1,T) = 1 - S(T),$$

and

$$S(T) = 1 - H_0(1,T) = 1 - \Gamma_0(H_1(1,T),T) = 1 - \Gamma_0(z_\infty(T),T),$$

where $z_\infty(T) = \Gamma_1(z_\infty(T),T) = H_1(1,T)$. Thus the size of the epidemic, if an epidemic occurs, is equal to the probability of an epidemic.

Compartmental models assume homogeneous mixing, corresponding to a Poisson network. As we shall see in the next section, for a Poisson network,

$$\Gamma_0(z,T) = \Gamma_1(z,T) = e^{\mathcal{R}_0(z-1)}.$$

Then the equation $\Gamma_1(z,T) = z$ is

$$e^{\mathcal{R}_0(z-1)} = z,$$

and the size of the epidemic is $1 - z_\infty(T)$. This is equivalent to the final size relation for a deterministic compartmental model [7, Sect. 1.3].

More sophisticated network analysis makes it possible to predict such quantities as the probability that an individual will set off an epidemic, the risk for an individual of becoming infected, the probability that a cluster of infections will set off a small disease outbreak when the transmissibility is less than the critical transmissibility, and how the probability of an epidemic depends on the degree of patient zero, the initial disease case [12, 14].

4.5 Some Examples of Contact Networks

The above analysis assumes that there is a known generating function $G_0(z)$ or, equivalently, a degree distribution $\{p_k\}$. In studying a disease outbreak, we need to know the degree distribution of the network. If we know the degree distribution we can calculate the basic reproduction number and also the probability of an epidemic. What kinds of networks are observed in practice in social interactions? There are some standard examples.

If contacts between members of the population are random, corresponding to the assumption of mass action in the transmission of disease, then the probabilities p_k are given by the *Poisson distribution*

$$p_k = \frac{e^{-c}c^k}{k!}$$

for some constant c. To show this, we think of a probability of contact $c\Delta t$ in a time interval Δt, and we let

$$n = \frac{1}{\Delta t}.$$

Then the probability of k contacts in a time interval Δt is

$$\binom{n}{k}(\frac{c}{n})^k(1-\frac{c}{n})^{n-k},$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the *binomial coefficient*. We rewrite this probability as

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{c^k}{k!}\frac{(1-\frac{c}{n})^n}{(1-\frac{c}{n})^k}.$$

We let $\Delta t \to 0$, or $n \to \infty$. Since

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \to 1, \quad (1-\frac{c}{n})^k \to 1,$$

and

$$(1-\frac{c}{n})^n \to e^{-c},$$

the limiting probability that there are k contacts is

$$p_k = \frac{e^{-c}c^k}{k!}.$$

Then the generating function is

$$G_0(z) = e^{-c}\sum_{k=0}^{\infty}\frac{c^k}{k!}z^k = e^{-c}e^{cz} = e^{c(z-1)} ,$$

and

$$G_0'(z) = ce^{c(z-1)}, \quad G_0'(1) = c.$$

The generating function for the Poisson distribution is $e^{c(z-1)}$. Then $G_1(z) = G_0(z)$, and $\mathcal{R}_0 = TG_1'(1) = cT$, so that

$$\Gamma_1(z,T) = G_1(1+(z-1)T) = e^{\mathcal{R}_0(z-1)}.$$

The commonly observed situation that most infectives do not pass on infection but there are a few "superspreading events" [17] corresponds to a probability distribution that is quite different from a Poisson distribution, and could give a quite different probability that an epidemic will occur. For example, taking $T = 1$ for simplicity, if $\mathcal{R}_0 = 2.5$ the assumption of a Poisson distribution gives $z_\infty = 0.107$ and $G_0(z_\infty) = 0.107$, so that the probability of an epidemic is 0.893. The assumption that nine out of ten infectives do not transmit infection while the tenth transmits 25 infections gives

$$G_0(z) = (z^{25}+9)/10, \quad G_1(z) = z^{24}, \quad z_\infty = 0, \quad G_0(z_\infty) = 0.9,$$

from which we see that the probability of an epidemic is 0.1. Another example, possibly more realistic, is to assume that a fraction $(1-p)$ of the population follows a Poisson distribution with constant r while the remaining fraction p consists of superspreaders each of whom makes L contacts. This would give a generating function

$$G_0(z) = (1-p)e^{r(z-1)} + pz^L$$
$$G_1(z) = \frac{r(1-p)e^{r(z-1)} + pLz^{L-1}}{r(1-p)+pL},$$

and

$$\mathcal{R}_0 = \frac{r^2(1-p)+pL(L-1)}{r(1-p)+pL}.$$

For example, if $r = 2.2, L = 10, p = 0.01$, numerical simulation gives

$$\mathcal{R}_0 = 2.5, \quad z_\infty = 0.146,$$

so that the probability of an epidemic is 0.849.

These examples demonstrate that the probability of a major epidemic depends strongly on the nature of the contact network. Simulations suggest that for a given value of the basic reproduction number the Poisson distribution is the one with the maximum probability of a major epidemic.

It has been observed that in many situations there is a small number of long range connections in the graph, allowing rapid spread of infection. There is a high degree of clustering (some vertices with many edges) and there are short path lengths. Such a situation may arise if a disease is spread to a distant location by an air traveller. This type of network is called a *small world* network. Long range connections in a network can increase the likelihood of an epidemic dramatically.

A third kind of network frequently observed is a *scale free* network. In a random network, the quantity p_k approaches zero very rapidly (exponentially) as $k \to \infty$. A scale free network has a "fatter tail", with p_k approaching zero as $k \to \infty$ but more slowly than in a random network. In an epidemic setting it corresponds to a situation in which there is an active core group but there are also "superspreaders" making many contacts. In a scale free network, p_k is proportional to $k^{-\alpha}$ with α a constant. In practice, α is usually between 2 and 3. Often an exponential cutoff is introduced in applications of scale free networks in order to make $G_0'(1) < \infty$ for every choice of α, so that

$$p_k = Ck^{-\alpha}e^{-k/\theta}.$$

The constant C, chosen so that $\sum_{k=0}^{\infty} p_k = 1$, can be expressed in terms of logarithmic integrals.

These examples indicate that the probability of an epidemic depends strongly on the contact network at the beginning of a disease outbreak. The study of complex networks is a field which is developing very rapidly. Some basic references are [15, 18], and other references to particular kinds of networks include [1, 2, 13, 19]. Examination of the contact network in a disease outbreak situation may lead to an estimate of the probability distribution for the number of contacts [11, 12], and thus to a prediction of the course of the disease outbreak.

A recent development in the study of networks in epidemic modeling is the construction of very detailed networks by observation of particular locations. The data that goes into such a network includes household sizes, age distributions, travel to schools, workplaces, and other public locations. The networks constructed are very complex but may offer a great deal of realism. However, it is very difficult to estimate how sensitive the predictions obtained from a model using such a complex network will be to small changes in the network. Nevertheless, simulations based on complicated networks are the

primary models currently being used for developing strategies to cope with a potential influenza pandemic. This approach has been followed in [8–10].

An alternative to simulations based on a very detailed network would be to analyze the behaviour of a model based on a simpler network, such as a random network or a scale-free network with parameters chosen to match the reproduction number corresponding to the detailed network. A truncated scale free network would have superspreaders and thus may be closer than a random network to what is often observed in actual epidemics.

4.6 Conclusions

We have described the beginning of a disease outbreak in terms of the degree distribution of a branching process, and have related this to a contact network. There is a developing theory of network epidemic models which is not confined to the early stages [12,14]. This involves more complicated considerations, such as the way in which a contact network may change over the course of an epidemic. We have restricted our attention to the beginning of an epidemic in order not to have to examine these complications. There are many aspects of network models for epidemics that have not yet been studied.

While we have suggested using a deterministic compartmental model once an epidemic is underway, it may be reasonable to go beyond the simplest Kermack–McKendrick epidemic model. Heterogeneity of contact rates, age structure, and other aspects of an actual epidemic can be modeled. Ideally, for the initial stages of an epidemic we would like to use a network somewhere between the over-simplification of a random network and the extreme complication of an individual-based model.

References

1. R. Albert & A.-L. Barabási: Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47–97 (2002).
2. A.-L. Barabási & R. Albert: Emergence of scaling in random networks, Science **286**, 509–512 (1999).
3. P. Erdös & A. Rényi: On random graphs, Publ. Math. **6**, 290–297 (1959).
4. P. Erdös & A. Rényi: On the evolution of random graphs, Pub. Math. Inst. Hung. Acad. Sci. **5**, 17–61 (1960).
5. P. Erdös & A. Rényi: On the strengths of connectedness of a random graph, Acta Math. Sci. Hung. **12**, 261–267 (1961).
6. D.S. Callaway, M.E.J. Newman, S.H. Strogatz & D.J. Watts: Network robustness and fragility: Percolation on random graphs, Phys. Rev. Lett. **85**, 5468–5471 (2000).
7. O. Diekmann & J.A.P. Heesterbeek: Mathematical Epidemiology of Infectious Diseases, Wiley, Chichester (2000).

8. N.M. Ferguson, D.A.T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn & D.S. Burke: Strategies for containing an emerging influenza pandemic in Southeast Asia, Nature **437**, 209–214 (2005).
9. I.M. Longini, M.E. Halloran, A. Nizam & Y. Yang: Containing pandemic influenza with antiviral agents, Am. J. Epidem. **159**, 623–633 (2004).
10. I.M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D.A.T. Cummings & M.E. Halloran: Containing pandemic influenza at the source, Science **309**, 1083–1087 (2005).
11. M.J. Keeling & K.T.D. Eames: Networks and epidemic models, J. R. Soc. Interface **2**, 295–307 (2006).
12. L.A. Meyers, B. Pourbohloul, M.E.J. Newman, D.M. Skowronski & R.C. Brunham: Network theory and SARS: Predicting outbreak diversity. J. Theor. Biol. **232**, 71–81 (2005).
13. A.L. Lloyd & R.M. May: Epidemiology: How viruses spread among computers and people, Science **292**, 1316–1317 (2001).
14. M.E.J. Newman: The spread of epidemic disease on networks, Phys. Rev. E **66**, 016128 (2002).
15. M.E.J. Newman: The structure and function of complex networks. SIAM Rev. **45**, 167–256 (2003).
16. M.E.J. Newman, S.H. Strogatz & D.J. Watts: Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E **64**, 026118 (2001).
17. S. Riley, C. Fraser, C.A. Donnelly, A.C. Ghani, L.J. Abu-Raddad, A.J. Hedley, G.M. Leung, L.-M. Ho, T.-H. Lam, T.Q. Thach, P. Chau, K.-P. Chan, S.-V. Lo, P.-Y. Leung, T. Tsang, W. Ho, K.-H. Lee, E.M.C. Lau, N.M. Ferguson & R.M. Anderson: Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions, Science **300**, 1961–1966 (2003).
18. S.H. Strogatz: Exploring complex networks, Nature **410**, 268–276 (2001).
19. D.J. Watts & S.H. Strogatz: Collective dynamics of 'small world' networks, Nature **393**, 440–442 (1998).