

9

Epidemic spreading in population networks

The mathematical modeling of epidemics is a very active field of research which crosses different disciplines. Epidemiologists, computer scientists, and social scientists share a common interest in studying spreading phenomena and rely on very similar models for the description of the diffusion of viruses, knowledge, and innovation. Epidemic modeling has developed an impressive array of methods and approaches aimed at describing various spreading phenomena, as well as incorporating many details affecting the spreading of real pathogens. In particular, understanding and predicting an epidemic outbreak requires a detailed knowledge of the contact networks defining the interactions between individuals. The theoretical framework for epidemic spreading has thus to be widened with opportune models and methods dealing with the intrinsic system complexity encountered in many real situations.

In this chapter, we introduce the general framework of epidemic modeling in complex networks, showing how the introduction of strong degree fluctuations leads to unusual results concerning the basic properties of disease spreading processes. Using some specific examples we show how plugging in complex networks in epidemic modeling enables one to obtain new interpretative frameworks for the spread of diseases and to provide a quantitative rationalization of general features observed in epidemic spreading. We end the chapter by discussing general issues about modeling the spread of diseases in complex environments, and we describe metapopulation models which are at the basis of modern computational epidemiology.

9.1 Epidemic models

The prediction of disease evolution and social contagion processes, can be conceptualized with a variety of mathematical models of spreading and diffusion processes. These models evolved from simple compartmental approaches into structured frameworks in which the hierarchies and heterogeneities present at the

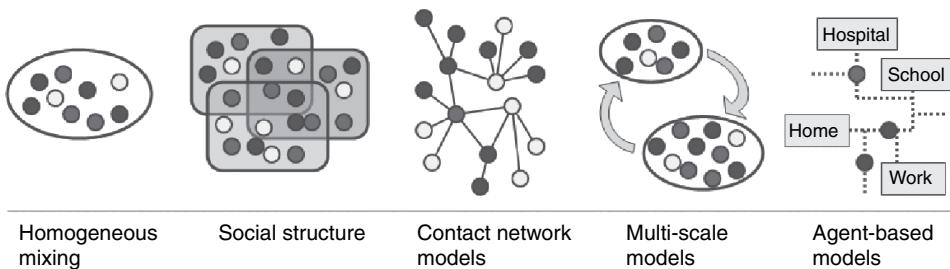


Fig. 9.1. Structures at different scales used in epidemic modeling. Circles represent individuals and each shade corresponds to a specific stage of the disease. From left to right: homogeneous mixing, in which individuals are assumed to interact homogeneously with each other at random; social structure, where people are classified according to demographic information (age, gender, etc.); contact network models, in which the detailed network of social interactions between individuals provide the possible virus propagation paths; multiscale models which consider subpopulations coupled by movements of individuals, while homogeneous mixing is assumed on the lower scale; agent-based models which recreate the movements and interactions of any single individual on a very detailed scale (a schematic representation of a part of a city is shown).

community and population levels are becoming increasingly important features (Anderson and May, 1992) (see Figure 9.1). As in many other fields, an interplay exists between the simplicity of the model and the accuracy of its predictions. An important question is how to optimize the predictive power of a model, which usually means reducing the number of tunable parameters. The challenging tasks of understanding and modeling the spread of diseases in countries, cities, etc., make this debate very real. At the geopolitical level, modeling approaches have evolved to explicitly include spatial structures and consist of multiple subpopulations coupled by movements among them, while the epidemics within each subpopulation are described according to approximations depending on the specific case studied (Hethcote, 1978; Anderson and May, 1984; May and Anderson, 1984; Bolker and Grenfell, 1993; Bolker and Grenfell, 1995; Lloyd and May, 1996; Grenfell and Bolker, 1998; Keeling and Rohani, 2002; Ferguson *et al.*, 2003). The modern versions of this patch or meta-population modeling scheme are multiscale frameworks in which different granularities of the system (country, inter-city, intra-city) are considered through different approximations and are coupled through interaction networks describing the flows of people and/or animals (Rvachev and Longini, 1985; Keeling *et al.*, 2001; Ferguson *et al.*, 2003; Hufnagel, Brockmann and Geisel, 2004; Longini *et al.*, 2005; Ferguson *et al.*, 2005; Colizza *et al.*, 2006a). Computers also serve as in-silico laboratories where the introduction of agent-based models (ABM) extends the modeling perspective, simulating propagation and contagion processes at the individual level (Chowell *et al.*, 2003; Eubank *et al.*, 2004).

Generally all these models are concerned with the evolution of the number and location of infected individuals in the population as a function of time. They therefore aim at understanding the properties of epidemics in the equilibrium or long time steady state, the existence of a non-zero density of infected individuals, the presence or absence of a global outbreak, the non-seasonal cycles that are observed in many infections, etc. (Anderson and May, 1992). A key parameter in the understanding of these properties is the basic reproductive number R_0 , which counts the number of secondary infected cases generated by one primary infected individual. It is easy to understand that any epidemic will spread across a non-zero fraction of the population only for $R_0 > 1$. In this case the epidemics are able to generate a number of infected individuals larger than those who have recovered, leading to an increase of the total number of infected individuals $I(t)$ at time t . This simple consideration leads to the definition of a crucial epidemiological concept, namely the epidemic threshold. Indeed, if the spreading rate is not large enough to allow a reproductive number larger than 1, the epidemic outbreak will not affect a finite portion of the population and dies out in a finite time. The epidemic threshold is a central concept in epidemiology since it provides a reference frame to understand the evolution of contagion processes and to focus on policies aimed at an effective reduction of the reproductive number in order to stop epidemic outbreaks.

9.1.1 Compartmental models and the homogeneous assumption

The simplest class of epidemic models assumes that the population can be divided into different classes or compartments depending on the stage of the disease (Anderson and May, 1992; Bailey, 1975; Daley and Gani, 2000; Diekmann and Heesterbeek, 2000), such as susceptibles (denoted by S , those who can contract the infection), infectious (I , those who have contracted the infection and are contagious), and recovered (R , those who have recovered from the disease). Additional compartments could be considered in order to model, for example, people immune to the disease or people exposed to the infection but not yet infectious (latent). According to this minimal framework, within each compartment individuals are assumed to be identical and homogeneously mixed, and the larger the number of sick and infectious individuals among one individual's contacts, the higher the probability of transmission of the infection. If we consider the spread of a disease in a population of N individuals, it is possible to obtain a general formulation of epidemic models which governs the time evolution of the number of individuals $X^{[m]}(t)$ in the class $[m]$ at time t (with $N = \sum_m X^{[m]}(t)$). The dynamics of the individuals between the different compartments depends on the specific disease considered. In general, the transition from a compartment to the other is specified by a reaction rate that depends on the disease etiology, such as the infection

transmission rate or the recovery or cure rate. In compartmental models there are two possible types of elementary processes ruling the disease dynamics. The first class of processes refers to the spontaneous transition of one individual from one compartment $[m]$ to another compartment $[h]$



Processes of this kind are the spontaneous recovery of infected individuals ($I \rightarrow R$) or the passage from a latent condition to an infectious one ($L \rightarrow I$) after the incubation period. In this case the variation in the number of individuals $X^{[m]}$ is simply given by $\sum_h v_h^m a_h X^{[h]}$, where a_h is the rate of transition from the class $[h]$ and $v_h^m = 1, 0$ or -1 is the change in the number of $X^{[m]}$ due to the spontaneous process from or to the compartment $[h]$.

The second class of processes refers to binary interactions among individuals such as the contagion of a susceptible individual in interaction with an infectious one



In this case, the variation of $X^{[m]}$ is given by $\sum_{h,g} v_{h,g}^m a_{h,g} N^{-1} X^{[h]} X^{[g]}$, where $a_{h,g}$ is the transition rate of the process and $v_{h,g}^m = 1, 0$ or -1 the change in the number of $X^{[m]}$ due to the interaction. The factor N^{-1} , where N is the number of individuals, stems from the fact that the above expression considers an homogeneous approximation in which the probability for each individual of class $[h]$ to interact with an individual of class $[g]$ is simply proportional to the density $X^{[g]}/N$ of such individuals. The homogeneous approximation is therefore equivalent to the mean-field one used for physical models and considers an effective interaction, a mass-action law, determining the force of infection in the same way for all individuals in the system. By using the above expressions it is possible to write the general deterministic reaction rate equations for the average number of individuals in the compartment $[m]$ as

$$\partial_t X^{[m]} = \sum_{h,g} v_{h,g}^m a_{h,g} N^{-1} X^{[h]} X^{[g]} + \sum_h v_h^m a_h X^{[h]}, \quad (9.4)$$

where now the quantities $X^{[m]}$ are continuous variables representing the average number of individuals in each class $[m]$. If the total number of individuals is constant, these equations must satisfy the conservation equation $\sum_m \partial_t X^{[m]} = 0$. It is also worth stressing that the deterministic continuous approximation neglects stochastic fluctuations that may be relevant in some cases, as we will see in the next sections.

This general framework easily allows us to derive the dynamical equations of the three basic models which are commonly used to illustrate the general properties of epidemic spreading processes. The simplest epidemiological model one can consider is the susceptible–infected (SI) model in which infected individuals can only exist in two discrete states, namely, susceptible and infected. The probability that a susceptible vertex acquires the infection from any given neighbor in an infinitesimal time interval dt is βdt , where β defines the pathogen *spreading rate*. Individuals that enter the infected class remain permanently infectious. The epidemics can only grow as the number of infectious individuals $I(t)$ is constantly increasing and the seed of infectious individuals placed at time $t = 0$ will therefore ultimately infect the rest of the population. The evolution of the SI model is therefore completely defined by the number of infected individuals $I(t)$ or equivalently the corresponding density $i(t) = I(t)/N$.

In the homogeneous assumption, the force of the infection (the per capita rate of acquisition of the disease for the susceptible individuals) is proportional to the average number of contacts with infected individuals, which for a total number of contacts k is approximated as ki . In a more microscopic perspective – which is at the basis of many numerical simulations of these processes – this can be understood by the following argument (see also Chapter 4). Since each infected individual attempts to infect a connected susceptible vertex with probability βdt , a susceptible vertex with n infected neighbors will have a total probability of getting infected during the time interval dt given by $1 - (1 - \beta dt)^n$. Neglecting fluctuations, each susceptible vertex with k connections will have on average $n = ki$ infected neighbors, yielding at the leading order in $\beta dt \ll 1$ an infection acquisition probability $1 - (1 - \beta dt)^{ki} \simeq \beta ki dt$ and the per capita acquisition rate βki .¹ This approach makes explicit the dependence of the spreading rate with the number of contacts k of each individual, which will be very useful later in extending the calculation in the case of heterogeneous systems. As a first approximation, let us consider that each individual/vertex has the same number of contacts/edges, $k \simeq \langle k \rangle$. The continuous and deterministic reaction rate equation describing the evolution of the SI model then reads as

$$\frac{di(t)}{dt} = \beta \langle k \rangle i(t) [1 - i(t)]. \quad (9.5)$$

The above equation states that the growth rate of infected individuals is proportional to the spreading rate $\beta \langle k \rangle$, the density of susceptible vertices that may become infected, $s(t) = 1 - i(t)$, where $s(t) = S(t)/N$, and the number of infected individuals in contact with any susceptible individual.

¹ Sometimes alternative definitions of the infection mechanism refer to βdt as the probability of acquiring the infection if one or more neighbors are infected. In this case the total acquisition probability is given by $\beta dt[1 - (1 - i)^k]$, i.e. the spreading probability times the probability that at least one neighbor is infected. Also in this case, for $\beta dt \ll 1$ and $i \ll 1$ an acquisition rate βki is recovered at the leading order.

The susceptible–infected–susceptible (SIS) model is mainly used as a paradigmatic model for the study of infectious diseases leading to an endemic state with a stationary and constant value for the prevalence of infected individuals, i.e. the degree to which the infection is widespread in the population as measured by the density of those infected. In the SIS model, individuals exist in the susceptible and infected classes only. The disease transmission is described as in the SI model, but infected individuals may recover and become susceptible again with probability μdt , where μ is the recovery rate. Individuals thus run stochastically through the cycle susceptible → infected → susceptible, hence the name of the model. The equation describing the evolution of the SIS model therefore contains a spontaneous transition term and reads as

$$\frac{di(t)}{dt} = -\mu i(t) + \beta \langle k \rangle i(t) [1 - i(t)]. \quad (9.6)$$

The usual normalization condition $s(t) = 1 - i(t)$ has to be valid at all times.

The SIS model does not take into account the possibility of an individual's removal through death or acquired immunization, which would lead to the so-called susceptible–infected–removed (SIR) model (Anderson and May, 1992; Murray, 2005). The SIR model, in fact, assumes that infected individuals disappear permanently from the network with rate μ and enter a new compartment R of removed individuals, whose density in the population is $r(t) = R(t)/N$. The introduction of a new compartment yields the following system of equations describing the dynamics:

$$\frac{ds(t)}{dt} = -\beta \langle k \rangle i(t) [1 - r(t) - i(t)] \quad (9.7)$$

$$\frac{di(t)}{dt} = -\mu i(t) + \beta \langle k \rangle i(t) [1 - r(t) - i(t)] \quad (9.7)$$

$$\frac{dr(t)}{dt} = \mu i(t). \quad (9.8)$$

Through these dynamics, all infected individuals will sooner or later enter the recovered compartment, so that it is clear that in the infinite time limit the epidemics must fade away. It is interesting to note that both the SIS and SIR models introduce a time scale $1/\mu$ governing the self-recovery of individuals. We can think of two extreme cases. If $1/\mu$ is smaller than the spreading time scale $1/\beta$, then the process is dominated by the natural recovery of infected to susceptible or removed individuals. This situation is less interesting since it corresponds to a dynamical process governed by the decay into a healthy state and the interaction with neighbors plays a minor role. The other extreme case is in the regime $1/\mu \gg 1/\beta$, i.e. a spreading time scale much smaller than the recovery time scale. In this case, as a first approximation, we can neglect the individual recovery that will

occur at a much later stage and focus on the early dynamics of the epidemic outbreak. This case corresponds to the simplified susceptible–infected (SI) model, for which infected nodes remain always infective and spread the infection to susceptible neighbors with rate β . In this limit the infection is bound to affect the whole population. The two limits therefore define two different regions in the parameters space and a transition from one regime to the other must occur at a particular value of the parameters, as we will see in the following.

9.1.2 The linear approximation and the epidemic threshold

All the basic models defined in the previous section can be easily solved at the early stage of the epidemics when we can assume that the number of infected individuals is a very small fraction of the total population; i.e. $i(t) \ll 1$. In this regime we can use a linear approximation for the equation governing the dynamics of infected individuals and neglect all i^2 terms. In the case of the SI model the resulting equation for the evolution of the density of infected individuals reads as

$$\frac{di(t)}{dt} = \beta \langle k \rangle i(t), \quad (9.9)$$

yielding the solution

$$i(t) \simeq i_0 e^{\beta \langle k \rangle t}, \quad (9.10)$$

where i_0 is the initial density of infected individuals. This solution simply states that the time scale $\tau = (\beta \langle k \rangle)^{-1}$ of the disease prevalence is inversely proportional to the spreading rate β ; the larger the spreading rate, the faster the outbreak will be. In the SI model the epidemic always propagates in the population until all individuals are infected, but the linear approximation breaks down when the density of infected individuals becomes appreciable and the behavior of $i(t)$ no longer follows a simple exponential form. For the SI model, it is in fact possible to derive the full solution of Equation (9.5), to obtain

$$i(t) = \frac{i_0 \exp(t/\tau)}{1 + i_0(\exp(t/\tau) - 1)}. \quad (9.11)$$

As expected, this expression recovers (9.10) for $t \ll \tau$, and shows saturation towards $i \rightarrow 1$ for $t \gg \tau$.

Along the same lines, it is possible to determine the early stage epidemic behavior of the SIS and SIR models. By neglecting the i^2 terms we obtain for both models the same linearized equation

$$\frac{di(t)}{dt} = -\mu i(t) + \beta \langle k \rangle i(t), \quad (9.12)$$

where in the case of the SIR model the term $r(t)$ can be considered of the same order as $i(t)$. The solution of the above equation is straightforward, yielding

$$i(t) \simeq i_0 e^{t/\tau}, \quad (9.13)$$

with i_0 the initial density of infected individuals and where τ is the typical outbreak time (Anderson and May, 1992)

$$\tau^{-1} = \beta \langle k \rangle - \mu. \quad (9.14)$$

The above relation contains a striking difference from the SI equation. The exponential characteristic time is the combination of two terms and can assume negative values if the recovery rate is large enough. In this case the epidemic will not spread across the population but will instead fade away on the time scale $|\tau|$. The previous consideration leads to the definition of a crucial epidemiological concept, namely the *epidemic threshold*. Indeed, if the spreading rate is not large enough (i.e., if $\beta < \mu/\langle k \rangle$), the epidemic outbreak will not affect a finite portion of the population and will die out in a finite time. The epidemic threshold condition can be readily written in the form

$$\tau^{-1} = \mu(R_0 - 1) > 0, \quad (9.15)$$

where $R_0 = \beta \langle k \rangle / \mu$ identifies the basic reproductive rate in the SIS and SIR models, which has to be larger than 1 for the spreading to occur.

The simple linear analysis allows the drawing of a general picture for epidemic evolution consisting of three basic stages (see Figure 9.2). Initially, when a few infected individuals are introduced into the population we define a pre-outbreak stage in which the evolution is noisy and dominated by the stochastic effects that are extremely relevant in the presence of a few contagious events. This stage is not described by the deterministic continuous equations derived in the homogeneous approximation and would require a full stochastic analysis. This is a stage in which epidemics may or may not disappear from the population just because of stochastic effects. When the infected individuals are enough to make stochastic effects negligible, but still very few if compared with the whole population, we enter the exponential spreading phase described by the linearized equations. In this case, the epidemic will pervade the system according to the exponential growth (9.13) whose time scale depends on the basic reproductive rate. Below the epidemic threshold the epidemic will just disappear in a finite time.

The final stage of the epidemic is model-dependent. The decrease of susceptible individuals reduces the force of infection of each infected individual (the i^2 term in Equation (9.5) is negative and therefore slows down the dynamics) and the exponential growth cannot be sustained any longer in the population. The SI model will continue its spread at a lower pace until the total population is in the infected state.

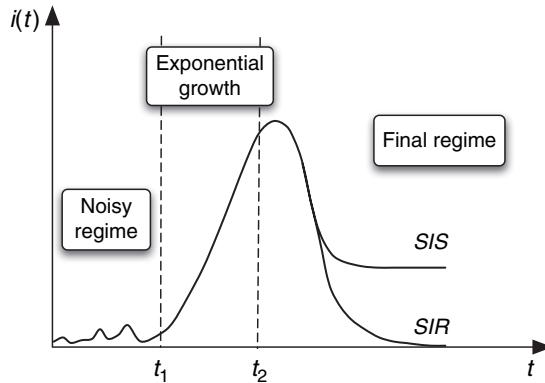


Fig. 9.2. Typical profile of the density of infected individuals $i(t)$ versus time on a given realization of the network. In the first regime $t < t_1$, the outbreak is subject to strong statistical fluctuations. In the second regime, $t_1 < t < t_2$ there is an exponential growth characterized by the reproductive number R_0 . In the final regime ($t > t_2$), the density of infected either converges to a constant for the SIS model or to zero for the SIR model. In individuals the SI model the epidemics will eventually pervade the whole population.

This cannot happen in the SIR and SIS model where individuals recover from the disease. The SIR model will inevitably enter a clean-up stage, since the susceptible compartment becomes depleted of individuals that flow into the removed compartment after the infectious period, and the epidemics will ultimately disappear. The SIS model will enter a stationary state in which the infectious individuals density is fixed by the balance of the spreading and recovery rate. However, it is worth stressing that while the outbreak will occur with finite probability if the parameters poise the system above the epidemic threshold, this probability is not equal to 1. Actually the stochastic fluctuations may lead to the extinction of the epidemics even well above the epidemic threshold. In general it is possible to estimate that the extinction probability of an epidemic starting with n infected individuals is equal to R_0^{-n} (Bailey, 1975). For instance, in the case of a single infected individual, even for values of R_0 as high as 2 the outbreak probability is just 50%.

The concept of epidemic threshold is very general and a key property of epidemic models. For instance, the addition of extra compartments such as latent or asymptomatic individuals defines models whose epidemic thresholds can still be calculated from the basic transition rates among compartments. Also, clustered but homogeneous connectivity patterns between individuals, such as regular lattices, meshes and even the Watts–Strogatz network, do not alter this scenario and just provide a different scaling behavior of the prevalence close to the threshold (Anderson and May, 1992; Marro and Dickman, 1999; Moore and Newman, 2000; Kuperman and Abramson, 2001). The relevance of the epidemic

threshold is also related to the protection of populations by means of immunization programs. These correspond to vaccination policies aimed at the eradication of the epidemics. The simplest immunization procedure one can consider consists of the random introduction of immune individuals in the population (Anderson and May, 1992), in order to get a *uniform* immunization density. In this case, for a fixed spreading rate β , the relevant control parameter is the density of immune vertices present in the network, the *immunity* g . At the mean-field level, the presence of a uniform immunity will have the effect of reducing the spreading rate β by a factor $1 - g$; indeed, the probability of infecting a susceptible and non-immune vertex will be $\beta(1 - g)[1 - i(t)]$. For homogeneous networks we can easily see that, for a constant β , we can define a critical immunization value g_c above which the effective system with spreading rate $\beta(1 - g)$ is pushed below the epidemic threshold:

$$g_c = 1 - \frac{\mu}{\beta \langle k \rangle}. \quad (9.16)$$

Thus, for a uniform immunization level larger than g_c , homogeneous networks are completely protected and no large epidemic outbreaks or endemic states are possible. The immunization threshold is very important in the prevention of epidemic outbreaks and in the clean-up stage. In practice, the aim of vaccination program deployment is to achieve a density of immunized individuals that pushes the population into the healthy region of the phase diagram.

9.2 Epidemics in heterogeneous networks

The general picture presented in the previous section is obtained in the framework of the homogeneous mixing hypothesis and by assuming that the network which describes the connectivity pattern among individuals is homogeneous; each individual in the system has, to a first approximation, the same number of connections $k \simeq \langle k \rangle$. However, many networks describing systems of epidemiological relevance exhibit a very heterogeneous topology. Indeed, social heterogeneity and the existence of “super-spreaders” have been known for a long time in the epidemics literature (Hethcote and Yorke, 1984). Similarly, much attention has been devoted to heterogeneous transmission rates and heterogeneous connectivity patterns. Generally, it is possible to show that the reproductive rate R_0 gets renormalized by fluctuations in the transmissibility or contact pattern as $R_0 \rightarrow R_0(1 + f(\sigma))$ where $f(\sigma)$ is a positive and increasing function of the standard deviation of the individual transmissibility or connectivity pattern (Anderson and May, 1992), pointing out that the spreading of epidemics may be favored in heterogeneous populations. Recently, empirical evidence has emphasized the role of heterogeneity by showing that many epidemiological networks are heavy-tailed and therefore the average

degree $\langle k \rangle$ is no longer the relevant variable. One then expects the fluctuations to play the main role in determining the epidemic properties. Examples of such networks relevant to epidemics studies include several mobility networks and the web of sexual contacts. Furthermore, computer virus spreading can be described in the same framework as biological epidemics (Kephart, White and Chess, 1993; Kephart *et al.*, 1997; Pastor-Satorras and Vespignani, 2001b; Aron *et al.*, 2001; Pastor-Satorras and Vespignani, 2004).

The presence of topological fluctuations virtually acting at all scales calls for a mathematical analysis where the degree variables explicitly enter the description of the system. This can be done by considering a degree block approximation that assumes that all nodes with the same degree are statistically equivalent. This assumption allows the grouping of nodes in the same degree class k , yielding the convenient representation of the system by quantities such as the density of infected nodes and susceptible nodes in the degree class k

$$i_k = \frac{I_k}{N_k}; \quad s_k = \frac{S_k}{N_k}, \quad (9.17)$$

where N_k is the number of nodes with degree k and I_k and S_k are the number of infected and susceptible nodes in that class, respectively. Clearly, the global averages are then given by the expressions

$$i = \sum_k P(k)i_k; \quad s = \sum_k P(k)s_k. \quad (9.18)$$

This formalism is extremely convenient in networks where the connectivity pattern dominates the system's behavior. When other attributes such as time or space become relevant, they must be added to the system's description: for instance, the time dependence is simply introduced through the time-dependent quantities $i_k(t)$ and $s_k(t)$.

9.2.1 The SI model

As a first assessment of the effect of the network heterogeneities, let us consider the simple SI model. In this case we know that the system is eventually totally infected whatever the spreading rate of the infection, but it is interesting to see the effect of topological fluctuations on the spreading velocity. In the case of the SI model the evolution equations read

$$\frac{di_k(t)}{dt} = \beta [1 - i_k(t)] k \Theta_k(t), \quad (9.19)$$

where the creation term is proportional to the spreading rate β , the degree k , the probability $1 - i_k$ that a vertex with degree k is not infected, and the density Θ_k

of infected neighbors of vertices of degree k . The latter term is thus the average probability that any given neighbor of a vertex of degree k is infected. This is a new and unknown quantity that in the homogeneous assumption is equal to the density of infected nodes. In a heterogeneous network, however, this is a complicated expression that must take into account the different degree classes and their connections.

The simplest situation we can face corresponds to a complete lack of degree correlations. As already discussed in the previous chapters, a network is said to have no degree correlations when the probability that an edge departing from a vertex of degree k arrives at a vertex of degree k' is independent from the degree of the initial vertex k . In this case, the conditional probability does not depend on the originating node and it is possible to show that $P(k'|k) = k'P(k')/\langle k \rangle$ (see Chapter 1 and Appendix 1). This relation simply states that any edge has a probability of pointing to a node with degree k' which is proportional to k' . By considering that at least one of the edges of each infected vertex points to another infected vertex from which the infection has been transmitted, one obtains

$$\Theta_k(t) = \Theta(t) = \frac{\sum_{k'} (k' - 1) P(k') i_{k'}(t)}{\langle k \rangle}, \quad (9.20)$$

where $\langle k \rangle = \sum_{k'} k' P(k')$ is the proper normalization factor dictated by the total number of edges. Thanks to the absence of correlations between the degrees of neighboring vertices, $\Theta_k(t)$ is then independent of k .

Combining Equations (9.19) and (9.20) one obtains the evolution equation for $\Theta(t)$. In the initial epidemic stages, neglecting terms of order $\mathcal{O}(i^2)$, the equations read

$$\frac{di_k(t)}{dt} = \beta k \Theta(t), \quad (9.21)$$

$$\frac{d\Theta(t)}{dt} = \beta \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) \Theta(t). \quad (9.22)$$

These equations can be solved and in the case of a uniform initial condition $i_k(t=0) = i_0$, the prevalence of nodes of degree k reads as

$$i_k(t) = i_0 \left[1 + \frac{k(\langle k \rangle - 1)}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau} - 1) \right], \quad (9.23)$$

with

$$\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}. \quad (9.24)$$

The prevalence therefore increases exponentially fast, and larger degree nodes display larger prevalence levels. The total average prevalence is also obtained as $i(t) = \sum_k P(k)i_k(t)$,

$$i(t) = i_0 \left[1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} (\mathrm{e}^{t/\tau} - 1) \right]. \quad (9.25)$$

The result (9.24) for uncorrelated networks implies that the growth time scale of an epidemic outbreak is related to the graph heterogeneity as measured by the heterogeneity ratio $\kappa = \langle k^2 \rangle / \langle k \rangle$ (see Chapter 2). In homogeneous networks with a Poisson degree distribution, in which $\kappa = \langle k \rangle + 1$, we recover the result $\tau = (\beta \langle k \rangle)^{-1}$, corresponding to the homogeneous mixing hypothesis. In networks with very heterogeneous connectivity patterns, on the other hand, κ is very large and the outbreak time scale τ is very small, signaling a very fast diffusion of the infection. In particular, in scale-free networks characterized by a degree exponent $2 < \gamma \leq 3$ we have that $\kappa \sim \langle k^2 \rangle \rightarrow \infty$ for networks of size $N \rightarrow \infty$. Therefore in uncorrelated scale-free networks we face a virtually instantaneous rise of the epidemic incidence. The physical reason is that once the disease has reached the hubs, it can spread very rapidly among the network following a “cascade” of decreasing degree classes (Barthélemy *et al.*, 2004; 2005).

9.2.2 The SIR and SIS models

The above results can be easily extended to the SIS and the SIR models. In the case of uncorrelated networks, Equation (9.19) contains, for both the SIS and the SIR models, an extra term $-\mu i_k(t)$ defining the rate at which infected individuals of degree k recover and become again susceptible or permanently immune (and thus removed from the population), respectively:

$$\frac{di_k(t)}{dt} = \beta k s_k(t) \Theta_k(t) - \mu i_k(t). \quad (9.26)$$

In the SIS model we have, as usual, $s_k(t) = 1 - i_k(t)$. In the SIR model, on the other hand, the normalization imposes that $s_k(t) = 1 - i_k(t) - r_k(t)$, where $r_k(t)$ is the density of removed individuals of degree k . The inclusion of the decaying term $-\mu i_k$, however, does not change the picture obtained in the SI model. By using the same approximations, the time scale for the SIR is found to behave as

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta) \langle k \rangle}. \quad (9.27)$$

In the case of diverging fluctuations the time scale behavior is therefore still dominated by $\langle k^2 \rangle$ and the spreading is faster for higher network heterogeneity. This

leads to a striking result concerning the epidemic threshold. In order to ensure an epidemic outbreak the basic condition $\tau > 0$ reads as²

$$\frac{\beta}{\mu} \geq \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \quad (9.28)$$

This implies that in heavy-tailed networks such that $\langle k^2 \rangle \rightarrow \infty$ in the limit of a network of infinite size we have *a null epidemic threshold*. While this is not the case in any finite size real-world network, larger heterogeneity levels lead to smaller epidemic thresholds. Also in this case, the parameter $\kappa = \langle k^2 \rangle / \langle k \rangle$ that defines the level of heterogeneity in the connectivity pattern is determining the properties of the physical processes occurring on the network. This is a very relevant result that, analogously to those concerning the resilience to damage (see Chapter 6), indicates that heterogeneous networks behave very differently from homogeneous networks with respect to physical and dynamical processes. In particular, the absence of any epidemic threshold makes scale-free networks a sort of ideal environment for the spreading of viruses, which even in the case of very weak spreading capabilities are able to pervade the network. We will see, however, that the picture is not so gloomy, as while the epidemic threshold is extremely reduced, the prevalence corresponding to small spreading rates is very small. Furthermore, we will see that the heterogeneity can be turned to our advantage by defining opportune vaccination strategies of great effectiveness.

9.2.3 The effect of mixing patterns

While we have so far restricted our study to the case of uncorrelated networks, it is worth noting that many real networks do not fulfil this assumption (Dorogovtsev and Mendes, 2003; Pastor-Satorras and Vespignani, 2004). In order to consider the presence of non-trivial correlations we have to fully take into account the structure of the conditional correlation function $P(k'|k)$. For the sake of simplicity let us consider the SI model. In this case, the equations we have written for the evolution of i_k can be stated as (Boguñá *et al.*, 2003b)

$$\begin{aligned} \frac{di_k(t)}{dt} &= \beta [1 - i_k(t)] k \Theta_k(t) \\ \Theta_k &= \sum_{k'} i_{k'} \frac{k' - 1}{k'} P(k'|k). \end{aligned} \quad (9.29)$$

² In the SIS model the equation is slightly different because in principle each infected vertex does not have to point to another infected vertex since the vertex from which it received the infection can spontaneously become susceptible again. This leads to $\Theta(t) = [\sum_{k'} k' P(k') i_{k'}(t)] / \langle k \rangle$, and consequently the epidemic threshold is given by $\beta/\mu \geq \langle k \rangle / \langle k^2 \rangle$.

The function Θ_k takes into account explicitly the structure of the conditional probability that an infected vertex with degree k' points to a vertex of degree k , with any of the $k' - 1$ free edges it has (not pointing to the original source of its infection). In the absence of correlations, $P(k'|k) = k'P(k')/\langle k \rangle$, recovering the results of Section 9.2.1. If the network presents correlations, measured by $P(k'|k)$, the situation is slightly more complex.

The evolution equation for $i_k(t)$ can be written at short times, neglecting terms of order $\mathcal{O}(i^2)$, as

$$\begin{aligned} \frac{di_k(t)}{dt} &= \sum_{k'} \beta k \frac{k' - 1}{k'} P(k'|k) i_{k'}(t) \\ &\equiv \sum_{k'} C_{k,k'} i_{k'}(t), \end{aligned} \quad (9.30)$$

which is a linear system of differential equations given by the matrix $\mathbf{C} = \{C_{k,k'}\}$ of elements

$$C_{k,k'} = \beta k \frac{k' - 1}{k'} P(k'|k). \quad (9.31)$$

Elementary considerations from mathematical analysis tell us that the behavior of $i_k(t)$ will be given by a linear combination of exponential functions of the form $\exp(\Lambda_i t)$, where Λ_i are the eigenvalues of the matrix \mathbf{C} . Therefore, the dominant behavior of the average prevalence will be

$$i(t) \sim e^{\Lambda_m t}, \quad (9.32)$$

where Λ_m is the largest eigenvalue of the matrix \mathbf{C} . The time scale governing the increase of the prevalence is thus given by $\tau \sim 1/\Lambda_m$. In the case of an uncorrelated network, the matrix \mathbf{C} , whose elements are $C_{k,k'}^{\text{nc}} = \beta k(k' - 1)P(k')/\langle k \rangle$, has a unique eigenvalue satisfying

$$\sum_{k'} C_{k,k'} \Psi_{k'} = \Lambda_m^{\text{nc}} \Psi_k, \quad (9.33)$$

where $\Lambda_m^{\text{nc}} = \beta(\langle k^2 \rangle / \langle k \rangle - 1)$, and where the corresponding eigenvector is $\Psi_k = k$, thus recovering the previous result Equation (9.24) of Section 9.2.1.

In the case of correlated networks, it has been shown using the Frobenius theorem (Gantmacher, 1974) that the largest eigenvalue of \mathbf{C} is bounded from below (Boguñá *et al.*, 2003)

$$\Lambda_m^2 \geq \min_k \sum_{k'} \sum_l (k' - 1)(l - 1) P(l|k) P(k'|l). \quad (9.34)$$

This equation is very interesting since it can be rewritten as

$$\Lambda_m^2 \geq \min_k \sum_l (l-1) P(l|k) (k_{\text{nn}}(l) - 1), \quad (9.35)$$

and it turns out (Boguñá *et al.*, 2003) that for scale-free networks with degree distribution $P(k) \sim k^{-\gamma}$ with $2 \leq \gamma \leq 3$, the average nearest neighbors degree $k_{\text{nn}}(l)$ diverges for infinite size systems ($N \rightarrow \infty$), which implies that Λ_m also diverges. Two particular cases have to be treated separately: it may happen that, for some k_0 , $P(l|k_0) = 0$; then the previous limit for Λ_m^2 gives no information but it is possible to show with slightly more involved calculations that Λ_m still diverges for $N \rightarrow \infty$ (Boguñá *et al.*, 2003b). Another problem arises if $k_{\text{nn}}(l)$ diverges only for $l = 1$; this only happens in particular networks where the singularity is accumulated in a pathological way onto vertices with a single edge. Explicit examples of this situation are provided by Moreno and Vázquez (2003).

The previous result, which has been also checked numerically (Moreno, Gómez and Pacheco, 2003a), has the important consequence that, even in the presence of correlations, the time scale $\tau \sim 1/\Lambda_m$ tends to zero in the thermodynamic limit for any scale-free network with $2 < \gamma \leq 3$. It also underlines the relevance of the quantity k_{nn} , which gives a lower bound for Λ_m in finite networks. This result can be generalized to both the SIR and SIS model, recovering the conclusion of the previous section concerning the epidemic time scale and the epidemic threshold. In the next sections we will analyze numerically both the SI and SIS models in order to provide a full description of the dynamical properties that takes into account the network's complexity as well as finite size effects in the population.

As we stated in Chapter 1, the correlation function $P(k|k')$ provides only a first account of the network structure. Important structural properties might be encoded in higher order statistical correlations and there is no complete theory of the effect of these properties on epidemic behavior. Recent works have confirmed the general picture concerning the behavior of the epidemic threshold in heterogeneous networks also in the case of non-trivial clustering properties (Serrano and Boguñá, 2006), but these general results may be altered by specific constructions of the network that break the statistical equivalence of nodes with respect to their degree.

9.2.4 Numerical simulations

In order to illustrate the analytical predictions presented in the previous sections, we show numerical simulations obtained by using an agent-based modeling strategy for an SI model in which at each time step the stochastic disease dynamics is applied to each vertex by considering the actual state of the vertex and its neighbors

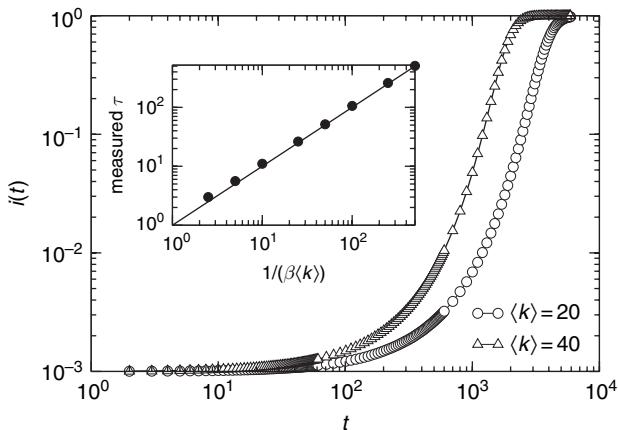


Fig. 9.3. Main frame: the symbols correspond to simulations of the SI model with $\beta = 10^{-4}$ on Erdős–Rényi networks with $N = 10^4$, $\langle k \rangle = 20, 40$; the lines are fits of the form of Equation (9.11). Inset: measured time scale τ , as obtained from fitting, versus the theoretical prediction for different values of $\langle k \rangle$ and β . From Barthélémy *et al.* (2005).

(see also Moreno, Gómez and Pacheco [2003a] for an alternative numerical method which directly solves the mean-field equations such as (9.29)). It is then possible to monitor the details of the spreading process, and to measure, for example, the evolution of the number of infected individuals. In addition, given the stochastic nature of the model, different initial conditions and network realizations can be used to obtain averaged quantities. We present simulations obtained for $N = 10^4$ and $\langle k \rangle$ ranging from 4 to 20. The results are typically averaged over a few hundred networks and for each network, over a few hundred different initial conditions. As an example of a homogeneous graph we consider the Erdős–Rényi network. In this case, Figure 9.3 shows the validity of Equation (9.11) and that the time scale of the exponential increase of the prevalence is given by $1/\beta\langle k \rangle$.

Needless to say, in the case of a homogeneous network, the hypothesis $k \simeq \langle k \rangle$ captures the correct dynamical behavior of the spreading. This is a standard result and we report the numerical simulations just as a reference for comparison with the following numerical experiments on heterogeneous networks. As a typical example of heterogeneous networks, the networks generated with the Barabási–Albert (BA) algorithm can be used (see Chapter 3). We report in Figure 9.4 the value of τ measured from the early time behavior of outbreaks in networks with different heterogeneity levels, as a function of $\langle k \rangle/\beta(\langle k^2 \rangle - \langle k \rangle)$. The numerical results recover the analytical prediction with great accuracy. Indeed, the BA network is a good example of uncorrelated heterogeneous networks in which the approximations used in the calculations are satisfied. In networks with correlations we expect

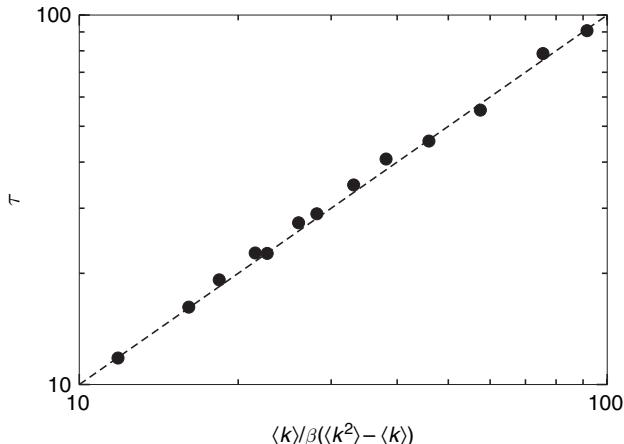


Fig. 9.4. Measured time scale τ in BA networks as obtained from exponential fitting to Equation (9.25) versus the theoretical prediction for different values of $\langle k \rangle$ and N corresponding to different levels of heterogeneity. From Barthélémy *et al.* (2005).

to find different quantitative results but qualitatively similar behavior, as happens in the case of the epidemic threshold evaluation (Boguñá *et al.*, 2003b).

While the present simulations are in very good agreement with the analytical results, the extent to which one expects such numerical simulations to recover exactly the analytical results is an open issue. The analytical results such as Equation (9.24) are obtained in the case of a linear expansion of the equations that is valid as long as $i(t) \ll 1$. On the other hand, the smaller the spreading time scale τ , the sooner the approximation breaks down. In very heterogeneous networks, the exponential regime may be extremely short or completely absent if the time scale of the epidemics is of the order of the unit time scale used in the simulations. In this case it is possible to show that the early time regime obeys a different behavior (Vázquez, 2006a; 2006c).

9.3 The large time limit of epidemic outbreaks

In the previous section we have limited ourselves to the analysis of the early time regime. It is, however, also interesting to study the opposite limit $t \rightarrow \infty$, in order to provide a more complete characterization of the epidemic outbreaks according to each model. The SI model is obviously uninteresting in the large time limit as it just asymptotically reaches the limit $i(t) \rightarrow 1$. The situation is very different if we consider the SIS and the SIR models. In these cases, the stationary prevalence or the total size of the epidemics depend upon the disease parameter of the model and the heterogeneity of the system.

9.3.1 The SIS model

The complete evolution equation for the SIS model on a network with arbitrary degree distribution can be written as

$$\frac{di_k(t)}{dt} = -\mu i_k(t) + \beta k [1 - i_k(t)] \Theta_k(t). \quad (9.36)$$

The creation term considers the density $1 - i_k(t)$ of susceptible vertices with k edges that might get the infection via a neighboring vertex. Let us first consider for the sake of simplicity the case in which the underlying network is a generalized random graph with no degree correlations. As already described in Sections 9.2.1 and 9.2.2, the calculation of Θ_k is then straightforward, as the average density of infected vertices pointed by any given edge that reads as

$$\Theta_k = \frac{1}{\langle k \rangle} \sum_{k'} k' P(k') i_{k'}, \quad (9.37)$$

which does not depend on k : $\Theta_k = \Theta$. Information on the $t \rightarrow \infty$ limit can be easily obtained by imposing the stationarity condition $di_k(t)/dt = 0$

$$i_k = \frac{k\beta\Theta}{\mu + k\beta\Theta}. \quad (9.38)$$

This set of equations shows that the higher a vertex degree, the higher its probability to be in an infected state. Injecting Equation (9.38) into (9.37), it is possible to obtain the self-consistent equation

$$\Theta = \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{\beta k \Theta}{\mu + \beta k \Theta}, \quad (9.39)$$

whose solution allows the calculation of Θ as a function of the disease parameters β and μ (Pastor-Satorras and Vespignani, 2001a; 2001b).

The epidemic threshold can be explicitly calculated from Equation (9.39) by just noting that the condition is given by the value of β and μ for which it is possible to obtain a non-zero solution Θ^* . Using a geometrical argument, as for the analysis of percolation theory in random graphs in Chapter 6, the solution of Equation (9.39) follows from the intersection of the curves $y_1(\Theta) = \Theta$ and $y_2(\Theta) = (1/\langle k \rangle) \sum_k k P(k) \beta k \Theta / (\mu + \beta k \Theta)$. The latter is a monotonously increasing function of Θ between the limits $y_2(0) = 0$ and $y_2(1) = (1/\langle k \rangle) \sum_k k P(k) \beta k / (\mu + \beta k) < 1$. In order to have a solution $\Theta^* \neq 0$, the slope of $y_2(\Theta)$ at the point $\Theta = 0$ must be larger than or equal to 1 (see Figure 9.5). This condition can be written as

$$\left. \frac{d}{d\Theta} \left(\frac{1}{\langle k \rangle} \sum_k k P(k) \frac{\beta k \Theta}{\mu + \beta k \Theta} \right) \right|_{\Theta=0} = \frac{\beta \langle k^2 \rangle}{\mu \langle k \rangle} \geq 1. \quad (9.40)$$

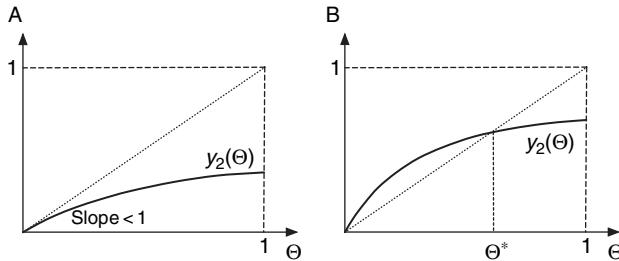


Fig. 9.5. Graphical solution of Equation (9.39). A, If the slope of the function $y_2(\Theta)$ at $\Theta = 0$ is smaller than 1, the only solution of the equation is $\Theta = 0$. B, When the slope is larger than 1, a non-trivial solution $\Theta^* \neq 0$ can be found.

The value of the disease parameters yielding the equality in Equation (9.40) defines the epidemic threshold condition that reads as

$$\frac{\beta}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (9.41)$$

This condition recovers the results obtained from the linear approximation at short times and confirms that topological fluctuations lower the epidemic threshold.

It is moreover possible to compute explicitly the behavior of the stationary density of infected individuals $i_\infty = \lim_{t \rightarrow \infty} i(t)$ as a function of the disease parameters in random uncorrelated scale-free networks where the heavy-tailed character is modeled by a power-law degree distribution with arbitrary exponent γ (Pastor-Satorras and Vespignani, 2001b). Consider a network which, in the continuous k approximation, has a normalized degree distribution $P(k) = (\gamma - 1)m^{\gamma-1}k^{-\gamma}$ and average degree $\langle k \rangle = (\gamma - 1)m/(\gamma - 2)$, where m is the minimum degree of any vertex. According to the general result (9.41), the epidemic threshold for infinite networks depends on the second moment of the degree distribution and is given, as a function of γ , by

$$\frac{\beta}{\mu} = \begin{cases} \frac{\gamma - 3}{m(\gamma - 2)} & \text{if } \gamma > 3 \\ 0 & \text{if } \gamma \leq 3 \end{cases}. \quad (9.42)$$

The behavior of the density of infected individuals in the stationary state may be found by solving explicitly the self-consistent equation for Θ in the limit of β/μ approaching the epidemic threshold. The full calculation yields different cases as a function of the exponent γ (Pastor-Satorras and Vespignani, 2001a):

(a) $2 < \gamma < 3$

In this case the leading order terms in β/μ of the solution yields

$$i_\infty \sim \left(\frac{\beta}{\mu} \right)^{1/(3-\gamma)}. \quad (9.43)$$

As expected, this relation does not show any epidemic threshold and gives a non-zero prevalence for all values of β/μ . It is important to note that the exponent governing the behavior of the prevalence, $1/(3 - \gamma)$, is larger than 1. This implies that for small β/μ the prevalence is growing very slowly, i.e. there exists a wide region of spreading rates in which $i_\infty \ll 1$.

(b) $\gamma = 3$

For this value of the degree exponent, logarithmic corrections dominate the scaling of the solution, yielding

$$i_\infty \sim e^{-\mu/m\beta}. \quad (9.44)$$

In this case, too, the absence of any epidemic threshold is recovered, and the prevalence approaches zero in a continuous way, exhibiting an exponentially small value for a wide range of spreading rates ($i_\infty \ll 1$).

(c) $3 < \gamma < 4$

The non-zero solution for Θ yields:

$$i_\infty \sim \left(\frac{\beta}{\mu} - \frac{\gamma - 3}{m(\gamma - 2)} \right)^{1/(\gamma-3)}. \quad (9.45)$$

That is, a power-law persistence behavior is observed. It is associated, however, to the presence of a non-zero threshold as given by Equation (9.42). Since $1/(\gamma-3) > 1$, the epidemic threshold is approached smoothly without any sign of the singular behavior associated to a critical point.

(d) $\gamma > 4$

The most relevant terms in the expansion of Θ now yield the behavior

$$i_\infty \sim \frac{\beta}{\mu} - \frac{\gamma - 3}{m(\gamma - 2)}. \quad (9.46)$$

That is, we recover the usual epidemic framework obtained for homogeneous networks.

In summary, the outcome of the analysis presented here is that the SIS model in scale-free uncorrelated random networks with degree exponent $\gamma \leq 3$ exhibits

the absence of an epidemic threshold or critical point. Only for $\gamma > 4$ do epidemics on scale-free networks have the same properties as on homogeneous networks.

9.3.2 The SIR model

In the case of the SIR model, the number of infected individuals is ultimately zero and the epidemics die because of the depletion of the susceptible individuals that after the infection move into the removed compartment. One of the main pieces of information on the course of the epidemics is therefore provided by the total number of individuals affected by the infection which is equal to the number of recovered individuals if the starting population was composed only of susceptible individuals. Taking into account the degree heterogeneity, this number is expressed as $r_\infty = \lim_{t \rightarrow \infty} r(t)$, where $r(t) = \sum_k P(k)r_k(t)$. This quantity may be explicitly calculated using the SIR equation for the degree classes (May and Lloyd, 2001; Moreno, Pastor-Satorras and Vespignani, 2002b; Newman, 2002b; Boguñá *et al.*, 2003b) which reads as

$$\frac{di_k(t)}{dt} = -\mu i_k(t) + \beta k s_k(t) \Theta_k(t), \quad (9.47)$$

$$\frac{ds_k(t)}{dt} = -\beta k s_k(t) \Theta_k(t), \quad (9.48)$$

$$\frac{dr_k(t)}{dt} = \mu i_k(t). \quad (9.49)$$

As in the previous sections, $\Theta_k(t)$ represents the average density of infected individuals at vertices pointed by any given edge and is given for an uncorrelated network by Equation (9.20). Equations (9.47), (9.48), (9.49), and (9.20), combined with the initial conditions $r_k(0) = 0$, $i_k(0)$, and $s_k(0) = 1 - i_k(0)$, completely define the SIR model on any random uncorrelated network with degree distribution $P(k)$. Let us consider the case of a homogeneous initial distribution of infected individuals, $i_k(0) = i_0$. In this case, in the limit of a very small number of initial infected individuals $i_0 \rightarrow 0$ and $s_k(0) \simeq 1$, Equations (9.48) and (9.49) can be directly integrated, yielding

$$s_k(t) = e^{-\beta k \phi(t)}, \quad r_k(t) = \mu \int_0^t i_k(\tau) d\tau, \quad (9.50)$$

where we have defined the auxiliary function

$$\phi(t) = \int_0^t \Theta(\tau) d\tau = \frac{1}{\langle k \rangle} \mu^{-1} \sum_k (k-1) P(k) r_k(t). \quad (9.51)$$

In order to obtain a closed relation for the total density of infected individuals, it is more convenient to focus on the time evolution of $\phi(t)$. To this purpose, let us compute its time derivative

$$\begin{aligned}\frac{d\phi(t)}{dt} &= \frac{1}{\langle k \rangle} \sum_k (k-1) P(k) i_k(t) \\ &= \frac{1}{\langle k \rangle} \sum_k (k-1) P(k) [1 - r_k(t) - s_k(t)] \\ &= 1 - \frac{1}{\langle k \rangle} - \mu \phi(t) - \frac{1}{\langle k \rangle} \sum_k (k-1) P(k) e^{-\beta k \phi(t)},\end{aligned}\quad (9.52)$$

where we have used the time dependence of $s_k(t)$ obtained in Equation (9.50). For a general distribution $P(k)$, Equation (9.52) cannot be solved in a closed form, but it is however possible to get useful information on the infinite time limit; i.e. at the end of the epidemics. In particular, the total epidemic prevalence $r_\infty = \sum_k P(k) r_k(\infty)$ is obtained as a function of $\phi_\infty = \lim_{t \rightarrow \infty} \phi(t)$

$$r_\infty = \sum_k P(k) (1 - e^{-\beta k \phi_\infty}), \quad (9.53)$$

where we have used $r_k(\infty) = 1 - s_k(\infty)$ and Equation (9.50). Since $i_k(\infty) = 0$, and consequently $\lim_{t \rightarrow \infty} d\phi(t)/dt = 0$, we obtain from Equation (9.52) the following self-consistent equation for ϕ_∞ :

$$\mu \phi_\infty = 1 - \frac{1}{\langle k \rangle} - \frac{1}{\langle k \rangle} \sum_k (k-1) P(k) e^{-\beta k \phi_\infty}. \quad (9.54)$$

The value $\phi_\infty = 0$ is always a solution. The non-zero ϕ_∞ solution, corresponding to finite prevalence $r_\infty > 0$, exists only if

$$\left. \frac{d}{d\phi_\infty} \left(1 - \frac{1}{\langle k \rangle} - \frac{1}{\langle k \rangle} \sum_k (k-1) P(k) e^{-\beta k \phi_\infty} \right) \right|_{\phi_\infty=0} \geq \mu. \quad (9.55)$$

This condition, which can be obtained with a geometrical argument analogous to that used to obtain Equation (9.40) for the SIS model, is equivalent to

$$\frac{\beta}{\langle k \rangle} \sum_k k(k-1) P(k) \geq \mu. \quad (9.56)$$

This defines the epidemic threshold condition

$$\frac{\beta}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \quad (9.57)$$

below which the epidemic prevalence is $r_\infty = 0$, and above which it reaches a finite value $r_\infty > 0$. This recovers the result (9.28) obtained by the linear approximation in the early stage of the dynamics and shows that the effects of topological fluctuations are consistently obtained in both limiting solutions.

The above equations for r_∞ cannot be solved explicitly for any network. As for the SIS model in the previous section, in the case of generalized random scale-free networks with degree distribution $P(k) = (\gamma - 1)m^{\gamma-1}k^{-\gamma}$ it is however possible to find an explicit solution that is an example of what happens in heavy-tailed networks (Moreno, Pastor-Satorras and Vespignani, 2002b; May and Lloyd, 2001; Newman, 2002b). Not surprisingly it is possible to show that r_∞ displays the same behavior as a function of β/μ as obtained for the stationary infected density i_∞ in the SIS model, and the various cases for the different ranges of γ values are recovered. In particular, for heavy-tailed networks of infinite size, and small values of β/μ , one obtains $r_\infty \sim (\beta/\mu)^{1/(3-\gamma)}$ and $r_\infty \sim e^{-\mu/m\beta}$ for $2 < \gamma < 3$ and $\gamma = 3$, respectively.

9.3.3 Epidemic models and phase transitions

The previous analysis allows us to draw an analogy between epidemic models and non-equilibrium continuous phase transitions. The SIS and SIR models are characterized by a threshold defining a transition between two very different regimes. These regimes are determined by the values of the disease parameters, and characterized by the global parameters i_∞ and r_∞ , which are zero below the threshold and assume a finite value above the threshold. From this perspective we can consider the epidemic threshold as the critical point of the system, and i_∞ and r_∞ represent the order parameter characterizing the transition in the SIS and SIR, respectively. Below the critical point the system relaxes in a frozen state with null dynamics, the healthy phase. Above this point, a dynamical state characterized by a macroscopic number of infected individuals sets in, defining an infected phase. Finally we have explained how, in the case of strong heterogeneity, the epidemic threshold is reduced and eventually suppressed by topological fluctuations in the case of infinite size networks.

It is therefore possible to draw a qualitative picture of the phase diagram of the system as depicted in Figure 9.6. The figure also shows the difference between homogeneous and heterogeneous networks where the epidemic threshold is shifted to very small values. On the other hand, for scale-free networks with degree distribution exponent $\gamma \leq 3$ the associated prevalence i_∞ is extremely small in a large region of values of β/μ . In other words the bad news of the suppression (or very small value) of the epidemic threshold is balanced by the very low prevalence attained by the epidemics.

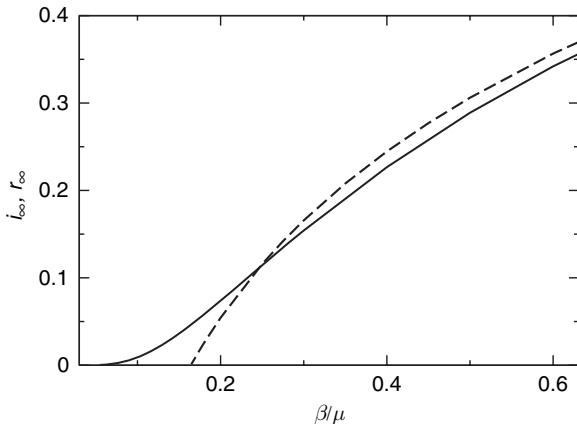


Fig. 9.6. Total prevalence (i_∞ for the SIS and r_∞ for the SIR model) in a heterogeneous network (full line) as a function of the spreading rate β/μ , compared with the theoretical prediction for a homogeneous network (dashed line). From Pastor-Satorras and Vespignani (2001a).

The mapping between epidemic models and non-equilibrium phase transitions has been pointed out in various contexts for a long time. As stressed by Grassberger (1983), the SIR model static properties can be mapped to an edge percolation process. Indeed, the epidemic threshold and the behavior close to threshold of the SIR model have the same form found for the percolation threshold in generalized networks (see Chapter 6). Analogously, it is possible to recognize that the SIS model is a generalization of the contact process model (Harris, 1974), widely studied as the paradigmatic example of an absorbing-state phase transition with a unique absorbing state (Marro and Dickman, 1999).

9.3.4 Finite size and correlations

As we discussed in the introductory chapters, real-world networks are composed of a number of elements that is generally far from the thermodynamic limit. This finite population introduces a maximum degree k_c , depending on the system size N or a finite connectivity capacity, which has the effect of restoring a bound to the degree fluctuations. The presence of the cut-off translates, through the general expression Equation (9.41), into an effective non-zero epidemic threshold due to *finite size effects*, as usually observed in non-equilibrium phase transitions (Pastor-Satorras and Vespignani, 2002a; May and Lloyd, 2001; Marro and Dickman, 1999). This positive epidemic threshold, however, is not an *intrinsic* property as in homogeneous systems, but an artifact of the limited system size that vanishes when increasing the network size or the degree cut-off.

To illustrate this point, let us focus on the SIS model in uncorrelated random networks with a scale-free degree distribution of the form $P(k) \simeq k^{-\gamma} \exp(-k/k_c)$. We define the epidemic threshold condition as

$$\frac{\beta}{\mu} = \rho_c, \quad (9.58)$$

and compute the effective nonzero epidemic threshold within the continuous k approximation (Pastor-Satorras and Vespignani, 2002a) as

$$\rho_c(k_c) = \frac{\langle k \rangle_{k_c}}{\langle k^2 \rangle_{k_c}} = \frac{\int_m^\infty k^{-\gamma+1} \exp(-k/k_c)}{\int_m^\infty k^{-\gamma+2} \exp(-k/k_c)} \equiv \frac{\Gamma(2-\gamma, m/k_c)}{\Gamma(3-\gamma, m/k_c)}. \quad (9.59)$$

Here m is the minimum degree of the network, and $\Gamma(x, y)$ is the incomplete Gamma function (Abramowitz and Stegun, 1972). For large k_c we can perform a Taylor expansion and retain only the leading term, obtaining for any $2 < \gamma < 3$ a threshold condition

$$\rho_c(k_c) \simeq \left(\frac{k_c}{m}\right)^{\gamma-3}. \quad (9.60)$$

The limit $\gamma \rightarrow 3$ corresponds instead to a logarithmic divergence, yielding at leading order

$$\rho_c(k_c) \simeq \frac{1}{m \ln(k_c/m)}. \quad (9.61)$$

It is interesting to compare the intrinsic epidemic threshold obtained in homogeneous networks with negligible degree fluctuations, $\rho_c^H = \langle k \rangle^{-1}$, with the non-zero effective threshold of bounded scale-free distributions. Figure 9.7 represents the ratio $\rho_c(k_c)/\rho_c^H$ as a function of k_c/m , for different values of the degree exponent γ . We can observe that, even for relatively small cut-offs ($k_c/m \sim 10^2 - 10^3$), for a reasonable value $\gamma \approx 2.5$ the effective epidemic threshold of finite scale-free networks is smaller by nearly an order of magnitude than the intrinsic threshold corresponding to a homogeneous network with the same average degree. This fact implies that the use of the homogeneity assumption would lead in scale-free networks to a serious over-estimation of the epidemic threshold, even for relatively small networks.

Another assumption that we have used in the calculation of the $t \rightarrow \infty$ properties of epidemic models is that the structure of the network is just characterized by its degree distribution. This is far from reality in most networks where structural properties and correlations are present. A full solution of the dynamical equations taking into account the general structural properties of the network is not possible, and we are left with approaches that progressively include higher order correlations. As we have seen for the early stage dynamics, a first step is

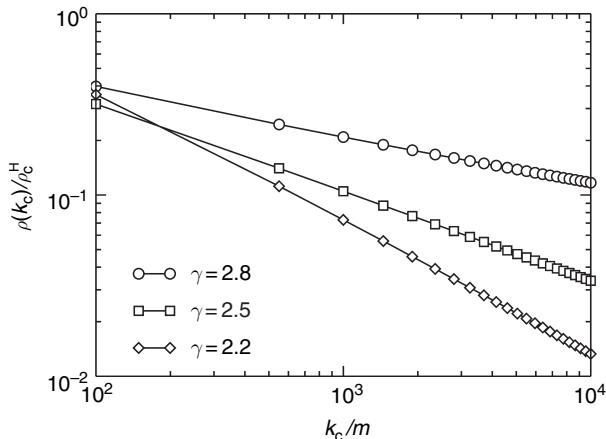


Fig. 9.7. Ratio between the effective epidemic threshold $\rho_c(k_c)$ in bounded scale-free networks with a soft exponential cut-off, and the intrinsic epidemic threshold ρ_c^H for homogeneous networks with the same average degree, for different values of γ . From Pastor-Satorras and Vespignani (2002a).

to include the correlations defined by the conditional probability $P(k' | k)$ that a vertex of degree k is connected to one of degree k' . Also, for the infinite time properties it is possible to show that the presence or lack of an epidemic threshold is directly related to the largest eigenvalue Λ_m of the connectivity matrix. This eigenvalue has been shown to diverge for all scale-free *unstructured networks* with infinite fluctuations,³ for any kind of two-point correlations (Boguñá and Pastor-Satorras, 2002; Boguñá *et al.*, 2003). Recent results (Serrano and Boguñá, 2006) have shown that more complicate structural properties, such as three vertices correlations and high clustering coefficient, do not alter the general framework concerning the epidemic threshold existence. On the other hand it is always possible to imagine very specific epidemic behaviors dictated by a peculiar construction or particular engineering of the network not completely captured by the statistical properties of the degree distribution and degree correlations.

Let us finally insist on the fact that the results presented in this chapter have been obtained within mean-field approximations. It is worth a word of warning that the domain of validity of these approximations is not a priori known. The numerical simulations have shown that this approach allows one to grasp and understand how the behavior of epidemic models is altered and strongly influenced by heterogeneous connectivity patterns. There is no guarantee that similar approximations will

³ The present result is only valid for networks with no internal structure, in which all the vertices with the same degree are statistically equivalent. It does not apply for regular lattices or structured networks (Klemm and Eguíluz, 2002b; Moreno and Vázquez, 2003), in which a spatial or class ordering constrains the connections among vertices.

be valid in the analytical study of any process taking place on networks. In particular, recent works have shown that mean-field predictions are contradicted by extensive numerical simulations in two particular models, namely the dynamical evolution of Ising spins located on the nodes of a network and interacting at zero temperature, and the contact process, in which particles located on the nodes of a network disappear or produce offspring according to particular rules (Castellano and Pastor-Satorras, 2006a; 2006b). It is therefore important to bear in mind that the mean-field approach, although usually very convenient and powerful, has to be systematically complemented by numerical investigations. The understanding of the limits of validity of the mean-field approximations represents, moreover, an interesting open problem.

9.4 Immunization of heterogeneous networks

The weakness of scale-free networks to epidemic attacks presents an extremely worrying scenario. The conceptual understanding of the mechanisms and causes for this weakness, however, allows us to develop new defensive strategies that take advantage of the heterogeneous topology. Thus, while random immunization strategies are utterly inefficient, it is yet possible to devise targeted immunization schemes which are extremely effective.

9.4.1 Uniform immunization

In heavy-tailed networks the introduction of a random immunization is able to depress the infection's prevalence locally, but it does so too slowly, being unable to find any critical fraction of immunized individuals that ensures the eradication of the infection. An intuitive argument showing the inadequacy of random immunization strategies is that they give the same importance to very connected vertices (with the largest infection potential) and to vertices with a very small degree. Because of the large fluctuations in the degree, heavily connected vertices, which are statistically very significant, can overcome the effect of the uniform immunization and maintain the endemic state.

In more mathematical terms, and as already stated in Section 9.1.2, the introduction of a density g of immune individuals chosen at random is equivalent to a simple rescaling of the effective spreading rate as $\beta \rightarrow \beta(1 - g)$, i.e. the rate at which new infected individuals appear is depressed by a factor proportional to the probability that they are not immunized. On the other hand, the absence of an epidemic threshold in the thermodynamic limit implies that any rescaling of the spreading rate does not bring the epidemic into the healthy region except in the case $g = 1$. Indeed, the immunization threshold g_c is obtained when the rescaled

spreading rate is set equal to the epidemic threshold. For instance, for uncorrelated networks we obtain

$$\frac{\beta}{\mu}(1 - g_c) = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (9.62)$$

In heavy-tailed networks with $\langle k^2 \rangle \rightarrow \infty$ only a complete immunization of the network ensures an infection-free stationary state in the thermodynamic limit (i.e. $g_c = 1$). The fact that uniform immunization strategies are less effective has been noted in the biological context in several cases of spatial heterogeneity (Anderson and May, 1992). In heavy-tailed networks, however, we face a limiting case due to the extremely high (virtually infinite) heterogeneity in the connectivity properties.

9.4.2 Targeted immunization

Although heavy-tailed networks hinder the efficiency of naive uniform immunization strategies, we can take advantage of their heterogeneity by devising immunization procedures that take into account the inherent hierarchy in the degree distribution. In fact, we know that heavy-tailed networks possess a noticeable resilience to *random* connection failures (Chapter 6), which implies that the network can resist a high level of accidental damage without losing its global connectivity properties, i.e. the possibility of finding a connected path between almost any two vertices in the system. At the same time, scale-free networks are strongly affected by *targeted* damage; if a few of the most connected vertices are removed, the network suffers a dramatic reduction of its ability to carry information. Applying this argument to the case of epidemic spreading, we can devise a *targeted* immunization scheme in which we progressively make immune the most highly connected vertices, which are the ones more likely to spread the disease. While this strategy is the simplest solution to the optimal immunization problem in heterogeneous populations (Anderson and May, 1992), its efficiency is comparable to the uniform strategies in homogeneous networks with finite degree variance. In heavy-tailed networks, it produces a striking increase of the network tolerance to infections at the price of a tiny fraction of immune individuals.

An approximate calculation of the immunization threshold in the case of a random scale-free network (Pastor-Satorras and Vespignani, 2002b) can be pursued along the lines of the analysis of the intentional attack of complex networks (see Section 6.5). Let us consider the situation in which a fraction g of the individuals with the highest degree have been successfully immunized. This corresponds, in the limit of a large network, to the introduction of an upper cut-off $k_c(g)$ – which is obviously a function of the immunization g – such that all vertices with

degree $k > k_c(g)$ are immune. At the same time, the infective agent cannot propagate along all the edges emanating from immune vertices, which translates into a probability $r(g)$ of deleting any individual contacts in the network. The elimination of edges and vertices for the spreading purposes yields a new connectivity pattern whose degree distribution and relative moments $\langle k \rangle_g$ and $\langle k^2 \rangle_g$ can be computed as a function of the density of immunized individuals (see the analogous calculation for the targeted removal of vertices in Section 6.5). The protection of the network will be achieved when the effective network on which the epidemic spreads satisfies the inequality $\langle k \rangle_g / \langle k^2 \rangle_g \geq \beta/\mu$, yielding the implicit equation for the immunization threshold

$$\frac{\langle k \rangle_{g_c}}{\langle k^2 \rangle_{g_c}} = \frac{\beta}{\mu}. \quad (9.63)$$

The immunization threshold is therefore an implicit function $g_c(\beta/\mu)$ and its analytic form will depend on the original degree distribution of the network.

In order to assess the efficiency of the targeted immunization scheme it is possible to perform the explicit calculation for an uncorrelated network with degree exponent $\gamma = 3$ (Pastor-Satorras and Vespignani, 2002b). In this case the leading order solution for the immunization threshold in the case of targeted immunization reads as

$$g_c \sim \exp(-2\mu/m\beta), \quad (9.64)$$

where m is the minimum degree of the network. This clearly indicates that the targeted immunization program is extremely convenient, with a critical immunization threshold that is exponentially small in a wide range of spreading rates. This theoretical prediction can be tested by performing direct numerical simulations of the SIS model on Barabási–Albert networks in the presence of targeted immunization. In Figure 9.8 the results of the targeted immunization are compared with simulations made with a uniform immunization (Pastor-Satorras and Vespignani, 2002b). The plot shows the reduced prevalence i_g/i_0 , where i_g is the stationary prevalence in the network with immunization density g and i_0 is the stationary prevalence in the non-immunized network, at a fixed spreading rate $\beta/\mu = 0.25$. This plot indicates that, for uniform immunization, the prevalence decays very slowly when increasing g , and will be effectively null only for $g \rightarrow 1$, as predicted by Equation (9.62).⁴ On the other hand, for the targeted immunization scheme, the prevalence shows a very sharp drop and exhibits the onset of a sharp immunization threshold above which the system is infection-free. A linear regression from the largest values of g yields an approximate immunization threshold

⁴ The threshold is not exactly 1 because of the usual finite size effects present even in the simulations which are performed on networks of size $N = 10^7$.

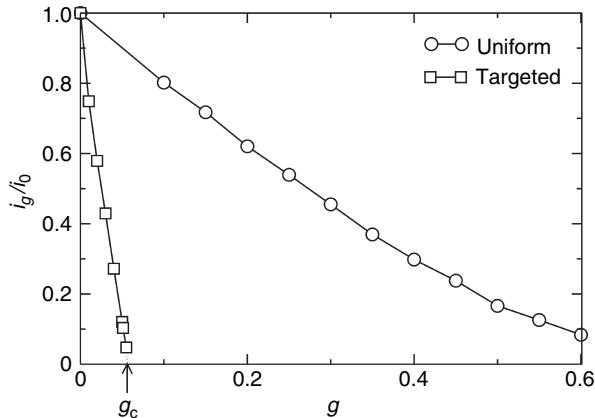


Fig. 9.8. Reduced prevalence i_g/i_0 from numerical simulations of the SIS model in the Barabási-Albert network (with $m = 2$) with uniform and targeted immunization, at a fixed spreading rate $\beta/\mu = 0.25$. A linear extrapolation from the largest values of g yields an estimate of the threshold $g_c \simeq 0.06$ for targeted immunization. From Pastor-Satorras and Vespignani (2002b).

$g_c \simeq 0.06$, that definitely proves that scale-free networks are very sensitive to the targeted immunization of a very small fraction of the most connected vertices.

Let us finally mention that, in a similar spirit, Dezsö and Barabási (2002) propose a level of safety and protection policy, which is proportional to the importance of the vertex measured as a function of its local degree. This implies that high degree vertices are cured with a rate proportional to their degree, or more generally to k^α . At the theoretical level it is possible to show that any $\alpha > 0$ reintroduces a finite epidemic threshold.

9.4.3 Immunization without global knowledge

While the targeted strategy is very effective, it suffers from a practical drawback in its real-world application. Its implementation requires a *complete* knowledge of the network structure in order to identify and immunize the most connected vertices. For this reason, several strategies to overcome this problem have been proposed, mainly relying just on a local, rather than a global, knowledge of the network. In particular, an ingenious immunization strategy was put forward by Cohen, Havlin and ben Avraham (2003), levering on a local exploration mechanism (see also Madar *et al.* [2004]). In this scheme, a fraction g of vertices are selected at random and each one is asked to point to one of its neighbors. The neighbors, rather than the selected vertices, are chosen for immunization. Since by following edges at random it is more probable to point to high degree vertices

which by definition have many links pointing to them, this strategy allows effective immunization of hubs without having any precise knowledge of the network connectivity. This strategy therefore manages to take advantage of the very same property that renders the network prone to infections. Variations on this idea have been subsequently proposed. A possibility consists in immunizing the vertex with highest degree found within shortest-path distance ℓ of randomly selected nodes (Gómez-Gardeñes, Echenique and Moreno, 2006). As ℓ increases, the necessary knowledge about the network's properties goes from local ($\ell = 1$) to global (when ℓ reaches the network's diameter). Holme (2004) also investigates *chained* versions of the immunization strategies, in which the nodes to be vaccinated are chosen successively as neighbors of previously vaccinated vertices. Such procedures turn out to be even more efficient, except in the case of very assortative or clustered networks. In the same spirit, it is possible to use the propagation properties of the heterogeneous networks to diffuse a vaccine on the network, starting from a random node, and using local heuristic propagation rules in which the probability that a node i sends the vaccine to a neighbor j increases with j 's degree and decreases with i 's, in order to preferentially immunize nodes with large degree (Stauffer and Barbosa, 2006).

In the particular case of the Internet, the propagation of viruses or worms is extremely fast. The development of vaccines occurs on a similar time scale, and dynamical and possibly collaborative response strategies have then to be developed in order to detect the virus and propagate the vaccine quickly.⁵ When viruses and antiviruses are competing on the same network, viruses have the inherent advantage of an earlier diffusion starting time. A way to mitigate this advantage consists of disseminating a random set of “honey-pots” in the network (the name honey-pot originates from their function as traps), and an extra set of special edges that can be traversed only by the immunizing agents and connect all honey-pots in a complete graph topology (Goldenberg *et al.*, 2005). Any virus spreading through the network rapidly reaches one of the honey-pots, triggering the defense mechanism: a vaccine is developed and spread immediately to all the other honey-pots, each of which acts as a seed for the propagation of the vaccine. This multiple propagation from the “superhub” formed by the clique of honey-pots transforms the vulnerability of scale-free networks into an advantage, since it uses both the possibility of a fast detection and the fast subsequent propagation. It could moreover be further improved by choosing particular nodes for their location, such as nodes obtained through the process of acquaintance immunization described above (Goldenberg *et al.*, 2005).

⁵ A detailed analysis of the computer science literature on this topic goes beyond the scope of the present book. (See Costa *et al.* [2005] for a recent work on such approaches.)

9.5 Complex networks and epidemic forecast

In recent years an impressive amount of study has focused on the effect of topological fluctuations on epidemic spreading. After the analysis of the SIS and SIR models, a long list of variations of the basic compartmental structure and the disease parameters have been proposed and studied in both the physics and mathematical epidemiology literature. While a detailed review of these papers goes beyond the scope of this book, a non-exhaustive list of recent works includes Sander *et al.* (2002); Joo and Lebowitz (2004); Dodds and Watts (2004); Olinky and Stone (2004); Petermann and De Los Rios (2004b); Dodds and Watts (2005); Crépey, Alvarez and Barthélémy (2006); Ahn *et al.* (2006); Karsai, Juhász and Iglói (2006); Vázquez (2006b); Zhou *et al.* (2006).

Most studies have mainly focused on systems in which each node of the network corresponds to a single individual; i.e. the network represents a single structured population. Only in recent years has the effect of heterogeneous connectivity patterns been studied for the case in which each node of the system may be occupied by any number of particles and the edges allow for the displacement of particles from one node to the other. Examples in which such a framework turns out to be relevant can be found in reaction–diffusion systems used to model a wide range of phenomena in chemistry and physics and mechanistic metapopulation epidemic models where particles represent people moving between different locations, such as cities or urban areas and the reaction processes between individuals simultaneously present at the same location are governed by the infection dynamics (Anderson and May, 1984; May and Anderson, 1984; Bolker and Grenfell, 1993; 1995; Lloyd and May, 1996; Grenfell and Bolker, 1998; Keeling and Rohani, 2002; Ferguson *et al.*, 2003; Watts *et al.*, 2005).

The development of such approaches and models, especially at the mechanistic level, is based on the detailed knowledge of the spatial structure of the environment, and of transportation infrastructures, movement patterns, and traffic networks. Most of these networks exhibit very heterogeneous topologies as in the case of the airport network among cities, the commuting patterns in inter- and intra-urban areas, and several info-structures (see Chapter 2). This clearly calls for a deeper understanding of the effect of heterogeneous connectivity patterns on these processes.

In more detail, metapopulation models describe spatially structured interacting subpopulations, such as cities, urban areas, or defined geographical regions (Hanski and Gaggiotti, 2004; Grenfell and Harwood, 1997). Individuals within each subpopulation are divided into the usual classes denoting their state with respect to the modeled disease (Anderson and May, 1992) and the compartment dynamics accounts for the possibility that individuals in the same location may get in contact

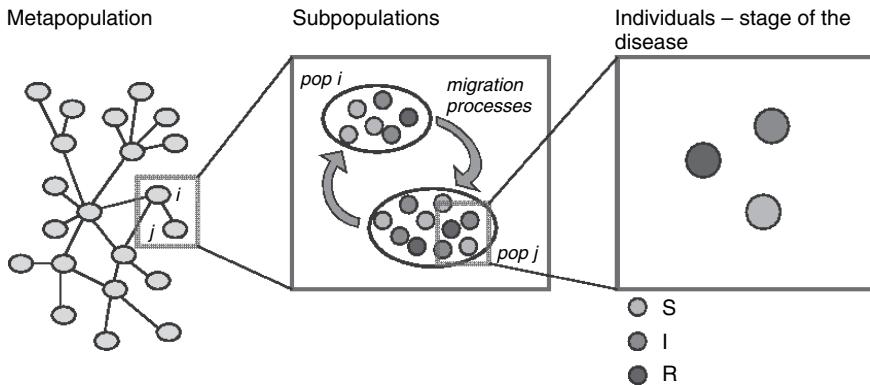


Fig. 9.9. Schematic representation of a metapopulation model. The system is composed of a heterogeneous network of subpopulations or patches, connected by migration processes. Each patch contains a population of individuals who are characterized with respect to their stage of the disease (e.g. susceptible, infected, removed), and identified with a different shade in the picture. Individuals can move from a subpopulation to another on the network of connections among subpopulations.

and change their state according to the infection dynamics. The interaction among subpopulations is the result of the movement of individuals from one subpopulation to the other (see Figure 9.9). A key issue in such a modeling approach lies in the accuracy with which we can describe the commuting patterns or traveling of people. In many instances even complicated mechanistic patterns can be accounted for by effective couplings expressed as a force of infection generated by the infectious individuals in subpopulation j on the individuals in subpopulation i (Bolker and Grenfell, 1995; Lloyd and May, 1996; Earn, Rohani and Grenfell, 1998; Rohani, Earn and Grenfell, 1999; Keeling, 2000; Park, Gubbins and Gilligan, 2002). An explicit mechanistic approach considers a detailed rate of traveling/commuting that defines the mixing subpopulation N_{ij} denoting the number of individuals of the subpopulation i present in the subpopulation j (Keeling and Rohani, 2002; Sattenspiel and Dietz, 1995). To these subpopulations are associated specific traveling and return rates that define different mobility patterns and stationary states, as in the work by Sattenspiel and Dietz (1995).

On the other hand, a simplified mechanistic approach uses a Markovian assumption in which at each time step the movement of individuals is given according to a matrix $p_{j\ell}$ that expresses the probability that an individual in the subpopulation j is traveling to the subpopulation ℓ . The Markovian character corresponds to the fact that we do not label individuals according to their original subpopulation (e.g. *home* in a commuting pattern framework) and at each time step the same traveling probability applies to all individuals in the subpopulation without having memory of their

origin. This approach is extensively used for very large populations when the traffic $w_{j\ell}$ among subpopulations is known, by stating that $p_{j\ell} \sim w_{j\ell}/N_j$ where N_j is the number of individuals in node j . Several modeling approaches to the large-scale spreading of infectious diseases use this mobility process based on transportation networks for which it is now possible to obtain detailed data (Baroyan *et al.*, 1969; Rvachev and Longini, 1985; Longini, 1988; Flahault and Valleron, 1991; Grais, Ellis and Glass, 2003; Grais *et al.*, 2004; Hufnagel *et al.*, 2004; Colizza *et al.*, 2007a; Colizza, Pastor-Satorras and Vespignani, 2007b).

In the case of a simple SIR model for the evolution of the disease, the metapopulation approach amounts to writing, for each subpopulation, equations such as (here for the variation of infected individuals)

$$\Delta I_j(t) = K(I_j, S_j, R_j) + \Omega_j(I), \quad (9.65)$$

where the first term of the r.h.s. of the equation represents the variation of infected individuals due to the infection dynamics within the subpopulation j and the second term corresponds to the net balance of infectious individuals traveling in and out of city j . This last term, the transport operator Ω_j , depends on the probability $p_{j\ell}$ that an infected individual will go from city j to city ℓ , and can be simply written as

$$\Omega_j(I) = \sum_{\ell} (p_{\ell j} I_{\ell} - p_{j\ell} I_j), \quad (9.66)$$

representing the total sum of infectious individuals arriving in the subpopulation j from all connected subpopulations ℓ , minus the amount of individuals traveling in the opposite directions. Similar equations can be written for all the compartments included in the disease model, finally leading to a set of equations where the transport operator acts as a coupling term among the evolution of the epidemics in the various subpopulations. It is clear that the previous equations represent a solid challenge to an analytical understanding. One way to tackle the issue is to look into simplified versions in which the coupling among the subpopulations is defined by a homogeneous diffusion rate p among all subpopulations linked in the network. The connectivity pattern of the metapopulation network is therefore simply described as a random graph characterized by a degree distribution $P(k)$ with given first and second moments $\langle k \rangle$ and $\langle k^2 \rangle$, respectively. By using a mechanistic approach it is possible to show that, along with the usual threshold condition $R_0 > 1$ on the local reproductive number within each subpopulation, the system exhibits a global invasion threshold that provides a critical value for the diffusion rate p below which the epidemics cannot propagate globally to a relevant fraction of subpopulations. Furthermore the global invasion threshold is affected by the topological fluctuations

of the metapopulation network. In particular, the larger the network heterogeneity, the smaller the value of the diffusion rate above which the epidemic may globally invade the metapopulation system. The relevance of these results stems from the fact that invasion thresholds typical of metapopulation models are also affected by heterogeneity and fluctuations in the underlying network, opening the path to a general discussion of threshold effects due to mobility fluxes in real-world networks.

Beyond the analysis of simplified metapopulation models, several approaches are focusing on data-driven implementation, aiming to explore the possibility of developing a computational infrastructure for reliable epidemic forecast and spreading scenario analysis. Taking into account the complexity of real systems in epidemic modeling has proved to be unavoidable, and the corresponding approaches have already produced a wealth of interesting results. But it is clear that many basic theoretical questions are still open. How does the complex nature of the real world affect our predictive capabilities in the realm of computational epidemiology? What are the fundamental limits on epidemic evolution predictability with computational modeling? How do they depend on the accuracy of our description and knowledge of the state of the system? Tackling such questions requires several techniques and approaches. Complex systems and networks analysis, mathematical biology, statistics, non-equilibrium statistical physics and computer science are all playing an important role in the development of a modern approach to computational epidemiology.