

Chapter 2

Percolation Methods for SEIR Epidemics on Graphs

Alberto Gandolfi

2.1 Introduction

Epidemic mathematical modelling has seen an enormous growth [32, 44, 4, 8], from the first deterministic models [50] to stochastic ones [3, 2, 20, 6], based on its ability to capture more and more features of real epidemics, building on the continuous development of mathematical techniques and computer power.

At about the same time percolation theory [77, 52, 42] has moved from simple [23] to more and more elaborate models [9] and detailed questions [27] providing a mathematically more developed theory than that of epidemics [83, 17].

Both percolation and most epidemic models are concerned with the spatial features of a random subset of some network, with the fundamental difference that while the spread of infectious diseases consists of a dynamical process, the mathematical theory of percolation is concerned with a static random object. Percolation is thus easier in principle and only particularly simple epidemic models are directly translated into percolation [41, 55]; however, it is becoming more and more clear over the years that there is in fact a more elaborate [73] and still mostly uncovered interplay between models in the two areas. This review plans to discuss some of the main features of this relation and to hint at some possible future developments.

There are a number of directions in which such interplay can be effective: as mentioned, in some simple cases or for some purposes, it is possible to disregard time in modelling an epidemic and be left with a percolation network. On other occasions, percolation-based observables can bound or approximate quantities of interest in an epidemic model. Again, it can be that the random set of infected individuals in an epidemic is itself modelled by the random set of a percolation

A. Gandolfi (✉)

Dipartimento di Matematica e Informatica U. Dini, Università di Firenze,
Viale Morgagni 67/A, 50134 Firenze, Italy
e-mail: gandolfi@math.unifi.it

process. Finally, since the networks on which epidemics take place can often be taken as random, the networks themselves could be built or analyzed by a suitably defined percolation process.

We review some examples in all these directions: the first three directions are explored in the first part of this review, while the second part deals with the last direction. The final part introduces a new epidemic model which incorporates most of the interesting features of epidemic modelling discussed in the other parts and which, due to its relation with various percolation models, is still suitable for explicit mathematical analysis. For simplicity and clarity we start the review from a specific stochastic model for an SEIR epidemic as discussed in the recent review [46], to which we refer for motivations and further details. For each example, we include a very simple calculation of the probability that one initial infected individual never transmits the infection in a network with four individuals and exponentially distributed random infectious times; the simple examples can be used to appreciate details and implementation difficulties; even at this simple level, models exhibit substantial differences and complications.

2.2 SEIR Epidemic Models and Percolation

2.2.1 *Deterministic Models*

Mathematical modelling of epidemics started with Daniel Bernoulli in the XIX century, and one of the major advances has been the 1927 Kermack–McKendrick model [50]. This describes an epidemic by the functions $s(t)$, $i(t)$ and $r(t)$ which denote the fractions of susceptible, infected, and recovered individuals at time t . Such functions must satisfy the condition $s(t) + i(t) + r(t) = 1$ and their evolution is described by the differential equations:

$$\begin{aligned} s'(t) &= -\lambda s(t)i(t) \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t) \\ r'(t) &= \gamma i(t) \end{aligned}$$

with starting points $s(0) = 1 - \epsilon$, $i(0) = \epsilon$ and $r(0) = 0$, where λ and γ are two parameters [20]. If $\lambda s(0) > \gamma$, then initially $i(t)$ is increasing (and only later decreasing) so that an outbreak is starting; otherwise, if $\lambda s(0) < \gamma$, then $i(t)$ decreases from the very start and infectives disappear steadily. The epidemic threshold is expressed in terms of the basic reproduction number $\mathbf{R}_0 = \lambda/\gamma$: for small ϵ , outbreak occurs at about $\mathbf{R}_0 = 1$. From the equations, \mathbf{R}_0 can be interpreted as the number of secondary infections from an infected individual. The final outbreak size of the epidemics $r(\infty)$ satisfies

$$1 - r(\infty) = (1 - \epsilon)e^{-\mathbf{R}_0 r(\infty)};$$

and one can see that $r(\infty)$ is then strictly positive only for $\mathbf{R}_0 > 1$ [20].

The Kermack–McKendrick model has many, often unrealistic, assumptions; in particular, the transmission of infectious diseases occurs at random. This led to stochastic models, on which we focus in the rest of the presentation.

2.2.2 General Stochastic SEIR Epidemic Model

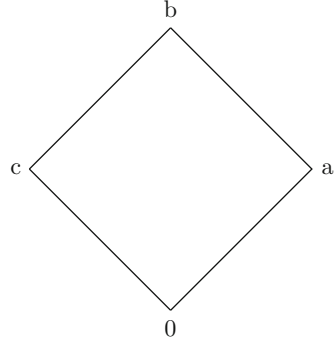
To introduce stochastic models, we start from a set V of individuals and a list of neighboring relations; it makes sense to take V finite, but properties which hold for large V are sometimes better understood in an infinite setting, so V can be countable as well. We thus have a reference network $G = (V, \mathcal{E})$, where $\mathcal{E} \subseteq \{(v_1, v_2), v_i \in V\}$ is the set of directed neighboring relations indicating that individual v_1 can possibly infect individual v_2 .

In a general stochastic SEIR model each individual $i \in V$ is in one of the four states: S (susceptible), E (exposed in the latent period), I (infected), and R (recovered or removed, thus immune). An individual $i \in V$ who is initially in state S is exposed at some random time t_i , where $t_i = \infty$ if i is never infected, moving to state E. After some random time e_i the individual i becomes infected, moving to state I, and after another random time h_i the individual moves to state R to signal perpetual recovery. The latent period e_i is a.s. finite and is identically 0 in the SIR model. The infectious period h_i is strictly positive and a.s. finite.

When i is in state I, j is in state S, and $(i, j) \in \mathcal{E}$ to signal network transmissibility, there is a random time $\tau_{i,j}^*$ such that at time $t_j = t_i + e_i + \tau_{i,j}^*$ the disease is transmitted to j who is thus exposed. Clearly, $\tau_{i,j}^* \in (0, h_i]$ or $\tau_{i,j}^* = \infty$ to indicate that there has been no transmission.

The starting state of the individuals is generally taken to be S for all but one or a few who start in state I, and the model is generally run by subsequent state changes: it can happen that a few individuals get exposed, indicating a minor outbreak, or that eventually a large proportion of individuals is infected and ends up in state R, signalling a major outbreak. More specifically, we have four mutually disjoint subsets $S_t, E_t, I_t, R_t \subseteq V$ which evolve for $t \in [0, \infty)$. In the initial state $E_0 = R_0 = \emptyset$ with S_0 and I_0 nonempty and often $|I_0| = 1$. Eventually, the evolution will reach a final stable state $S_\infty, E_\infty = \emptyset, I_\infty = \emptyset, R_\infty$. The model is thus identified by the reference graph G and the distributions $F_i^E(e)$, $F_i^I(h)$ and $F_{i,j}^*(\tau|h)$ of e_i , h_i and $\tau_{i,j}^*$, respectively, where the last one indicates a conditional distribution given the value of h [46]. The major issue in epidemic analysis is then concerned with the proportion of R_∞ which for large networks is generally either very small or considerably large; a large outbreak is defined as the case $R_\infty \gg 0$, for instance equal to a positive fraction of $|V|$. Other features are related to the parameters of the above distributions, such as the parameters range in which a large outbreak can take place, the probability of such an outbreak, the expected final size of the infected population and the size distribution, the transmission probabilities to a

Fig. 2.1 The elementary network used for some simple examples



given individual, the size distribution of infected group in case of no major outbreak, and the time evolution.

Many epidemic models can be retrieved from the general stochastic SEIR model by appropriate choices of the network G and of the distributions which identify the model. Although the model is complicated by the simultaneous presence of a large number of random variables, its analysis can be quite often simplified. In the opposite direction, it will be convenient to add additional random variables to capture more features of real epidemics, as we discuss later. The next section starts with some simplifications.

Here is the first of a series of simple examples. In all examples $V = \{0, a, b, c\}$ and $I_0 = \{0\}$, and we compute the probability $\text{Prob}(R_\infty = \{0\})$ that the epidemics does not spread from 0. The only relevant infectious period is that of 0, which we take to have rate μ : $F_0^H(h) = 1 - e^{-\mu h}$. After the neighboring relations \mathcal{E} of the network are determined, for each $i, j \in V$, $(i, j) \in \mathcal{E}$, the transmission rate once i is infected is some $\beta_{i,j}$ so $F_{i,j}^*(\tau|h) = 1 - e^{-\beta_{i,j}h\tau}$. One can easily see that the latent periods e_i 's do not play any role in the calculations.

Example 1 (General stochastic SEIR epidemics on a network Fig. 2.1). Let $\mathcal{E} = \{\{0, a\}, \{0, b\}, \{a, c\}, \{b, c\}\}$. Then

$$\text{Prob}(R_\infty = \{0\}) = \int_0^\infty \mu e^{-\mu h} e^{-h(\beta_{0,a} + \beta_{0,b})} dh = \frac{\mu}{\mu + \beta_{0,a} + \beta_{0,b}}$$

2.2.3 Time Evolution

Although epidemics have a time development which is clearly of interest and could be studied from the model [14], many issues can be more easily studied by reducing or eliminating the time dependence of the model. We start our simplification from here.

Already, by the definition of the general stochastic SEIR model one can run the model in two equivalent ways: “on the fly,” which mimics the evolution of the epidemics and amounts to determining the value of the random variables when needed; or “a priori” assigning all the values of the random variables before determining the epidemics and then using values if and when needed [46]. In the “a priori” method one samples for each individual i the variables e_i and h_i from their respective distributions, and then for each ordered pair (i, j) one samples $\tau_{i,j}^*$ from the conditional distribution $F_{i,j}^*(\tau | h)$ for the already sampled $h = h_i$.

The a priori sampling suggests that, if we aim at disregarding time, all we really care is whether $\tau_{i,j}^* \leq h_i$ or $\tau_{i,j}^* = \infty$, in other words whether there has been transmission of the disease from i to j or not. In fact, if $\tau_{i,j}^* \leq h_i$ and i has been infected, then the infection reaches j : either it reaches it from the infectious contact with i , or it has already reached it from a previous infectious contact. In any case, $t_j < \infty$ and the spread can continue from j . Such information is sufficient to determine all the features related to R_∞ , although not those related to R_t and the t_i 's.

So we can introduce the random probability $p(i, j, h_i) = \text{Prob}(\tau_{i,j}^* \neq \infty)$ and discard both the random times $\tau_{i,j}^*$ and e_i from the model. After this, the only remaining random variables are the h_i 's.

Example 2 (Timeless general stochastic SEIR epidemics on a network). Let \mathcal{E} be as in Example 1. The only difference is the introduction of $p(i, j, h_i) = 1 - e^{-\beta_{i,j}h_i}$ so

$$\text{Prob}(R_\infty = \{0\}) = \int_0^\infty \mu e^{-\mu h} (1 - p(0, a, h))(1 - p(0, b, h)) dh = \frac{\mu}{\mu + \beta_{0,a} + \beta_{0,b}}$$

If the infectious periods are constant $h_i \equiv h$ and the probabilities $p(i, j, h_i) = p(i, j, h) = p(i, j) = p$ do not depend on i and j , time can be discretized and we get the classical Reed–Frost model [8]. The simplification is enormous, and unrealistic, but clearly one can get more explicit results [57, 43]. The model is then equivalent to the most elementary percolation process.

2.2.4 Bernoulli Percolation

Mathematical percolation models have been devised as extremely simplified models of some random medium still capable of capturing features like sudden global changes as consequence of particular microscopic small changes. For instance, one can think of the passage of fluid in a porous medium: mathematical percolation would model the medium as a random subset of some regular networks with edges open to the fluid independently selected: with the increasing density of open edges the system remains impermeable till it suddenly becomes almost completely permeable.

In more formal mathematical detail, the *standard Bernoulli bond percolation* with parameter $p \in [0, 1]$ on a graph $G = (V, \mathcal{E})$ starts from independently and

Fig. 2.2 A percolation cluster in \mathbb{Z}^2



randomly selecting each edge $e \in \mathcal{E}$ with probability p and then studying the behavior of the generated random subset formed by the vertices V and the selected edges. One often focuses on maximal connected components of the random subgraph called clusters. In *site percolation*, instead, the vertices are randomly selected and the random subgraph is formed by the selected vertices and the edges incident to them. Bond percolation is used to model porous media, interactions, spread of fire, and so on, while site percolation can model randomly distributed particles [23], mixed materials, etc. [52, 77]. The qualitative properties of the two models are generally the same if the dimension of the space in which they are embedded is the same, but quantitative features vary.

Although the definition is in general terms, the theory becomes interesting when G is embedded in some geometric space, for instance is the d -dimensional integer lattice or some other two-dimensional lattice (Fig. 2.2). In fact, the relative simplicity of the process allows to deal with the constraints of a spatial structure, a relevant feature also for epidemics which is ignored in the simplest models. This is the main reason why percolation theory can be useful in studying epidemics.

When G is infinite, one key question in percolation theory is whether there are, with positive probability, infinite clusters. One of the first results in percolation theory is that, depending on the graph, there exists a critical $p_c \in (0, 1)$ such that for $p < p_c$ there are no infinite clusters and for $p > p_c$ there is an infinite cluster containing a fixed given vertex (the origin, for instance) with positive probability: in this case we say that percolation occurs. In \mathbb{Z}^2 , $p_c = 1/2$ [51] for bond percolation and it is an unknown value estimated around 0.59 for site

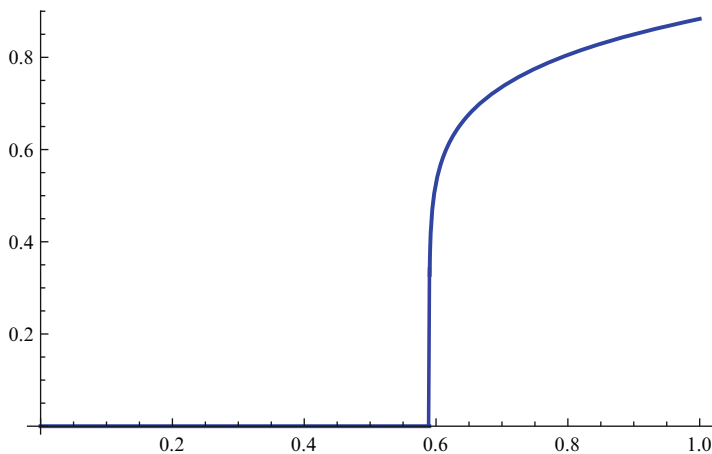


Fig. 2.3 Two-dimensional site percolation probability

percolation [42]. Moreover, the transition is sharp, in the sense that the probability that the origin is in an infinite cluster grows very rapidly to one (with a jump in such a function or in its derivative) (Fig. 2.3).

Example 3 (Bernoulli percolation on a network). Let \mathcal{E} be as in Example 1. In Bernoulli percolation with parameter p

$$\text{Prob}(\text{ the origin } 0 \text{ is isolated }) = (1 - p)^2.$$

Other quantities studied in percolation theory have a corresponding meaning in the epidemic model. The measure of the development of the epidemics by the successive generations, which is to say that for each infected site we record the minimum number of infections needed to reach it, is called chemical distance and the diffusion of the outbreak measured by the chemical distance growths linearly with time instead of exponentially as it does in the general epidemic model. In particular, a shape theorem holds, in the sense that denoting by A_n the set of vertices infected using at most n infections, there exists a deterministic set $A \subseteq \mathbb{Z}^d$ such that $A_n/n \rightarrow A$ with probability one (see [40]). In addition, if percolation occurs the infinite cluster is unique, and, on the other hand, the sets of infected individuals in the absence of a big outbreak, which is finite clusters in the subcritical regime, have an exponentially small size [42].

In percolation theory, critical exponents describe features like the shape of the percolation probability, or the distribution of the size of the cluster of the origin when $p = p_c$. The exponents originate by the fact that it is believed that several quantities are polynomial, with a critical exponent as power. For instance, the probability θ_p that the origin is in an infinite cluster is believed to behave like $\theta_p \approx (p - p_c)^\beta$. The value of the critical exponents is believed to depend only on essential features of the graph, like dimensionality, but not on local details, defining

some universality classes of models which share the same critical exponents [17]. A similar theory is also expected to hold for epidemic models: [65, 41, 28] argue that epidemics are governed by the “dynamics percolation” universality class.

A mathematically very challenging theory has been developed to describe the asymptotic behavior of large clusters at the critical point p_c [27, 83], showing that several quantities and random subsets become conformal invariant in the scaling limit (i.e., the network mesh going to zero). Results so far hold only for very specific two-dimensional graphs [76, 26].

Clearly, one can consider random edges up to some bounded length: this is called *short-range Bernoulli percolation* and basically no new phenomena appear.

2.2.5 Reed–Frost Model

In the general stochastic SEIR model, if $p(i, j, h_i) = p(i, j)$, then each directed edge between pairs of individuals is generated independently since the only dependencies between different transmissions from a vertex i stem from the common parameter h_i . In the Reed–Frost model $p(i, j) = p \in [0, 1]$. Transmissions are then independent and identically distributed (i.i.d.) random variables, and the “a priori” realization is exactly the Bernoulli percolation model with parameter p : in fact, the presence of an edge $\{i, j\}$ in the percolation configuration corresponds to the transmission through that edge when the first of i and j is exposed to the infection; if none is ever exposed, then the presence of $\{i, j\}$ can be disregarded, and if $\{i, j\}$ is not present and the other individual is later exposed, then transmission through $\{i, j\}$ is in any case irrelevant to the future course of the epidemics.

We then have a complete transcription: the set of infected individuals is the percolation cluster of the origin, the distribution of small epidemics is the distribution of subcritical clusters, the critical value for epidemics is the percolation critical point, the probability of outbreak is the probability of an infinite percolation cluster, the final size of the epidemic is the density of the infinite cluster, and the probability of being infected is the probability of being in the cluster of the origin.

Although this is an unrealistic simplification, it can serve as a benchmark or for testing plausible features of the model. When we later discuss realistic networks, the direct translation of the Reed–Frost epidemic model into Bernoulli percolation can be used to get a first insight into the features of some classes of networks.

Example 4 (Comparison of general stochastic SEIR epidemics with percolation). Let \mathcal{E} be as in Example 1 and the general SEIR epidemics. It is possible to obtain a Bernoulli percolation, or, equivalently, a Reed–Frost epidemic model, by averaging the infectious periods $\bar{h} \equiv \int_0^\infty h\mu e^{\mu h} dh = 1/\mu$ and then taking $p(\bar{h}) = 1 - e^{-\frac{\beta_{0,a} + \beta_{0,b}}{2} \frac{1}{\mu}}$.

If the transmission rates $\beta_{i,j} = \beta$ do not depend on the individuals, then one can also take the average transmission probability $\bar{p} = \bar{p}(0, a) = \bar{p}(0, b) = \int_0^\infty$

$(1 - e^{-\beta})\mu e^{\mu h} dh = \frac{\beta}{\mu + \beta}$ and consider transmissions to be independent, as in Example 3.

Then

$$\begin{aligned} \text{Prob}(R_\infty = \{0\}) &= (1 - p(\bar{h}))^2 = e^{-(\beta_{0,a} + \beta_{0,b})\frac{1}{\mu}}, \\ \text{Prob}'(R_\infty = \{0\}) &= \left(1 - \frac{\beta}{\mu + \beta}\right)^2 = \left(\frac{\mu}{\mu + \beta}\right)^2. \end{aligned}$$

Since $e^{-(\beta_{0,a} + \beta_{0,b})\frac{1}{\mu}} < \left(\frac{\mu}{\mu + 2\beta}\right)$ and $\left(\frac{\mu}{\mu + \beta}\right)^2 < \left(\frac{\mu}{\mu + 2\beta}\right)$ for small μ , percolation with average infectious periods (see Sect. 2.2.10 below for a development of this bound) or that with average transmission probabilities overestimate the outbreak probability computed in Example 1. None of the two reproduces the exact quantity.

2.2.6 Basic Reproduction Number

With the translation of Reed–Frost epidemic into percolation one begins to discover how the spread of infectious diseases takes place between individuals dispersed in a space with some geometrical features.

The first discovery is that the value $\mathbf{R}_0 = 1$ of the basic reproduction number does not, in general, identify the threshold. For instance, in \mathbb{Z}^2 , the percolation critical point is $p_c = 1/2$, but at $p = 1/2$ we have $\mathbf{R}_0 = 2$. The reason is that saturation occurs quite rapidly and infected individuals are surrounded by already recovered individuals, so that a larger value of \mathbf{R}_0 is needed. On the other hand, in many epidemics the value of \mathbf{R}_0 greatly exceeds 1.

Since the basic reproduction number \mathbf{R}_0 is often available in concrete situations, it can be used to estimate the model parameters from which the occurrence of percolation can be inferred, often by means of simulations. In the Reed–Frost model on \mathbb{Z}^d , $p = \mathbf{R}_0/4$ and thus a large epidemic occurs if $\mathbf{R}_0/4 > p_c = 1/2$.

2.2.7 Fixed Infectious Periods

When the infectious periods $h_i \equiv h$ are constant and the probabilities $p(i, j, h_i) = p(i, j)$ then it is still possible to formulate a percolation network equivalent to the general stochastic SEIR model. First generate i.i.d. random variables $p(i, j)$'s for each ordered pair (i, j) of individuals. Suppose that the epidemics starts at some vertex (the origin, for instance) and let $\{i, j\} \in E$; once again, if i is the first to be exposed to the infectious disease between those at i and j , then use the realization of $p(i, j)$; if the first exposed is j , then use the realization of $p(j, i)$, and if none among i and j is infected, then take one of the realizations of $p(i, j)$ or $p(j, i)$ at random. As

we argued before, possible transmission from the second individual exposed has no relevance to the course of the epidemic. In this construction, the edges have been taken independently, thus realizing a (nonhomogeneous) percolation process. Notice that this occurs in spite of the fact that the $p(i, j)$'s depend on the directed edges: this is because we really care only about the first infected in the pair of neighbors and on whether it transmits the disease or not; the second infected and its attempt to transmit are irrelevant in the SIR model. Independence derives from the fact that the $p(i, j)$'s are independent for different edges, and the infectious period is constant. Then, in the SIR epidemic with constant infection times also we have a complete transcription in terms of percolation.

Example 5 (Inhomogeneous Bernoulli percolation on a network). Let E be as in Example 1. With $p(i, j)$ as above we have

$$\text{Prob}(\text{the origin } 0 \text{ is isolated}) = (1 - p(0, a))(1 - p(0, b)).$$

2.2.8 Plague Among Great Gerbils in Kazakhstan

A more direct relation to percolation has been discovered recently in the interpretation of the threshold for the plague epidemic among great gerbils in Kazakhstan: in [33] data about a plague epidemic (infection with *Yersinia pestis*) among great gerbils in Kazakhstan has been analyzed. Great gerbils build a burrow system with a small group living in each burrow and data recorded by monitoring in Kazakhstan shows a threshold related to the abundance (fraction of occupied burrows). In [33] a realistic network is taken, with appropriate occupied burrow spatial density, and then an SIR epidemic is simulated by assuming that one burrow is infected and that at each infected site at exponential times an attempted infection occurs; the destination of the infectious attempt has been based on the study of great gerbil movements which are responsible for carrying fleas around: the destination burrow of the infectious attempt is then determined with a random distribution based on the distance from the infected burrow. Outbreak is defined as the spread of infection at some distance. As the observation distance is increased, the probability of outbreak goes from linearly increasing in the burrow occupation rate to a discontinuous function jumping at some critical value. The simulations indicate a threshold at burrow occupation rate $p \approx 0.31$, while the collected data indicated a threshold at a 95 % confidence interval of about (0.287, 0.373). At the same time the basic reproduction number at the threshold has been estimated as $R_0 \approx 1.5$. Thus it appears that a percolation phase transition is taking place even among a non-static population.

Notice that the explanatory model consists of some modification of the standard percolation models: burrows are not placed at the vertices of a regular two-dimensional network; in addition, both site variables (in terms of occupied burrows) and bond variables (in terms of transmissions) are present; and, finally,

connections are not only between closely vertices, but can also reach over a long distance. This last remark brings us to long-range percolation, which is discussed in a later section.

2.2.9 Epidemic Percolation

Back to the general stochastic SEIR model: it can be given a graphical representation using directed edges. Once the random variables h_i and $p(i, j, h_i)$ are determined, for each i and j in V if $(i, j) \in \mathcal{E}$ we select the directed edge (i, j) with probability $p(i, j, h_i)$. Then, starting from the set I_0 of initially infected individuals we determine R_∞ by including I_0 and all the individuals who can be reached from I_0 following the directed edges selected above.

For a given I_0 , the random set R_∞ incorporates all of the relevant information about the epidemics except time evolution. Such a random set is in fact a percolation process, named epidemics percolation [46]. Differently from the standard percolation processes, the edges are directed, but more importantly, dependent.

The directionality of the model implies that it has to be analyzed by considering the outcomponent and incomponent of each individual i , which is to say, the sets of individuals which are reachable from i in the forward direction, and the set of individuals from which i can be reached [46].

The dependence comes from the common parameter h_i used to determine the $p(i, j, h_i)$'s with common i [55], as in Example 2. The selection of the directed edges (i, j) has, however, only a local dependence: in fact, if $i \neq i'$, then (i, j) and (i', j') are independently selected. Such condition is called one-dependence, since there is dependence only up to distance one [39]. One-dependence implies that if $\beta_{i,j} = \beta$ for all pairs of individuals, then there is a critical β_c such that for $\beta < \beta_c$ the probability P_β of a big (infinite) outbreak is zero, while $P_\beta > 0$ for $\beta > \beta_c$ [81, 9]. Bernoulli percolation with parameter equal to the average transmission probabilities overestimate the large outbreak probability [54], as we have seen in Example 4.

From a general point of view, the fact that we now relate epidemics to directed percolation suggests that this is in a different universality class than that related to ordinary percolation.

2.2.10 Extra Variables: Infectivity and Susceptibility

Additional features of the spread of infectious diseases can be incorporated in the stochastic SEIR model by additional individual-based random variables. Infectivity W_i and susceptibility \bar{W}_i are introduced in [59] as i.i.d. random variables, not

necessarily independent for fixed i , and letting $p(i, j) = p(W_i, \bar{W}_j)$ be the transmission probability from i to j once the first is infected and the second susceptible.

This model deals with different aspects of the infectious transmission, but in some instances it is equivalent to the general stochastic SEIR model: if $p(i, j, h_i) = p(h_i)$ depends on i only through h_i and does not depend on j , then the SEIR model is equivalent to the Meester–Trapman model with $W_i = h_i, \bar{W}_i = 1$ and the same function p .

The main result of [59] is that under general conditions several features of epidemics such as the probability of a large outbreak and the expected final size of the outbreak are bounded by the corresponding percolation quantities (probability of percolation and expected cluster size) from above by a Bernoulli bond percolation model (obtained by letting $p(x, y) = x y$, $W_i \equiv W$, $\bar{W}_i \equiv \bar{W}$, $W\bar{W} = p$, with W and \bar{W} constants) and from below by a Bernoulli site percolation model (obtained by letting $p(x, y) = x y$ and $p = P(W_i = \bar{W}_i = 1) = 1 - P(W_i = \bar{W}_i = 0)$); we indicate the Bernoulli bond percolation model by P_p^{bond} and the Bernoulli site percolation model by P_p^{site} . In particular, suppose that for some random variables W and \bar{W} , $W_i \equiv W$ and $\bar{W}_i \equiv \bar{W}$, then if Ξ is a collection of hopable paths (i.e., a collection closed under the operation of switching paths at crossing points and which can be suitably approximated by the first parts of the paths), then the event C^Ξ that for at least one path in Ξ all of its edges are present satisfies the following [59]. Let (W, \bar{W}) be a random vector and let $p(x, y) = p(x y)$ be such that $p(z)$ is increasing and concave. Then, for every

$$p \geq p(\max[E(W\bar{W}), E(W)E(\bar{W})]),$$

we have

$$P(C^\Xi) \leq P_p^{bond}(C^\Xi).$$

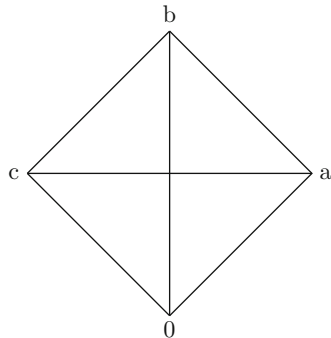
A similar result shows that if $p \leq E(W\bar{W})$ then $P(C^\Xi) \geq P_p^{site}(C^\Xi)$. These results allow us to estimate the probability of a large outbreak, and hence the epidemics critical threshold, and the probability that a given vertex will ever be infected, by corresponding quantities in the percolation model, which are generally relatively easier to compute or estimate.

2.3 Random Networks and Percolation

2.3.1 Random Graphs

So far we have paid little attention to the network structure. The Kermack–McKendrick model can be recovered from the general SEIR epidemic model when G is the complete graph. Models on complete graph $\mathcal{E} = V^2$ are called mass-action or mean-field models.

Fig. 2.4 The complete network



Example 6 (Kermack–McKendrick model [50] Fig. 2.4). Let $\mathcal{E} = V \times V$ and let the transmission rates be constant. In this simple example we omit the usual rescaling by the inverse of the population size; we then have $\beta_{i,j} \equiv \beta$. Then

$$\text{Prob}(R_\infty = \{0\}) = \int_0^\infty \mu e^{-\mu h} (1 - p(h))^3 dh = \frac{\mu}{\mu + 3\beta}.$$

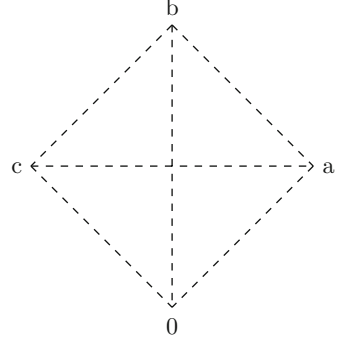
As we mentioned, mass-action models do not capture the inhomogeneity in the number of individuals that can be potentially infected by a given individual: this is better described with a reference network, in particular if this is random.

The classic example of a random network [5] is the Bernoulli random graph, usually called Erdős–Rényi [36, 37]. Starting from n vertices the random graph $G(n, p)$ is generated by connecting each pair of vertices independently with probability p . Depending on the value of p there are different asymptotic behaviors of the random graph as n diverges and one important transition is called the emergence of the giant component. If $np < 1$, then a graph in $G(n, p)$ will almost surely have no connected components of size larger than $O(\log n)$. If $np = 1$, then a graph in $G(n, p)$ will almost surely have largest component whose size is of order $n^{2/3}$. If np tends to a constant $c > 1$, then a graph in $G(n, p)$ will almost surely have a unique giant component containing a positive fraction of the vertices and no other component will contain more than $O(\log n)$ vertices.

Using the random graph G as a reference graph for the standard stochastic SEIR model, a large outbreak is possible only for $np \geq 1$. Due to the independence of the edges in the random graph, this model actually produces just a change of transmission probability. This is seen by a joint construction of epidemic evolution and the random graph [63]. With the usual initial sets S_0 and I_0 , the evolution is split into successive discretized generations identified by S_k, I_k, R_k . Each individual in I_k makes an independent infectious attempt on each individual in S_k which is successful with probability $p(h_i)$; if the infection is transmitted by at least one successful attempt, then the target individual enters I_{k+1} , else it goes into S_{k+1} . Finally, $R_{k+1} = I_k \cup R_k$ and $S_{k+1} = S_k \setminus I_{k+1}$.

The timeless stochastic SEIR epidemic evolution is correctly reconstructed by this procedure, and one can retrieve the random graph as well. For any edge $\{i, j\}$, if

Fig. 2.5 The basic random network; dashed lines are random



$i, j \notin R_\infty$, then the edge is taken with probability p ; otherwise, if the infectious attempt through $\{i, j\}$ has been successful (in whichever direction it was attempted), then $\{i, j\}$ is taken; if the unique infection attempt was not successful (and was attempted from i), then $\{i, j\}$ is taken with probability $\frac{p_n(1-p(h_i))}{1-p_n p(h_i)}$ [63]. Then, any edge through which a transmission has been attempted or realized is taken with probability $p_n p(h_i) + (1 - p_n p(h_i)) \frac{p_n(1-p(h_i))}{1-p_n p(h_i)} = p$ and all edges are selected independently.

Asymptotics for large n are discussed in [64] and [63], where it is shown that in this case if $p = p_n = \alpha/n$ for some $\alpha > 0$ and the transmission rates are rescaled by the average number $np_n - 1$ of neighbors in the random graph so that $\beta_{i,j} \equiv \beta / (np_n - 1)$, then the basic reproduction number is $\mathbf{R}_0 = \alpha \int_0^\infty \mu e^{-\mu h} e^{-\frac{\beta}{np_n - 1} h} dh$. Note that \mathbf{R}_0 is the way to identify the epidemic threshold since these are mass-action models.

Example 7 (SEIR on a random graph Fig. 2.5). The network edges in \mathcal{E} are chosen randomly with uniform probability p_4 ; transmission probabilities should be rescaled by the factor $4p_4 - 1$ but we avoid it in this example to simplify comparison with the other examples. Then $p(i, j, h) = 1 - e^{-\beta h}$. Hence

$$\begin{aligned}
 \text{Prob}(R_\infty = \{0\}) &= \int_0^\infty \mu e^{-\mu h} \sum_{k=0}^3 \binom{3}{k} p_4^k (1-p_4)^{3-k} (1-p(h))^k dh \\
 &= \int_0^\infty \mu e^{-\mu h} (1-p_4 p(h))^3 dh \\
 &= \sum_{k=0}^3 \binom{3}{k} p_4^k (1-p_4)^{3-k} \int_0^\infty \mu e^{-\mu h} e^{-\beta k h} dh \\
 &= \sum_{k=0}^3 \binom{3}{k} p_4^k (1-p_4)^{3-k} \frac{\mu}{\mu + \beta k}.
 \end{aligned}$$

In this case, the basic reproduction number satisfies $\mathbf{R}_0 = \frac{\alpha \mu}{\beta + \mu}$.

2.3.2 Small-World Networks

In spite of the interesting developments of the last section, random graphs are not satisfactory from some points of view. Starting from a paper by Watts and Strogatz [82] new indicators have been used to classify graphs and new classes of networks have been discovered. One indicator is the average point-to-point distance L on the graph, obtained by averaging, over all pairs of vertices on the graph, their graph distance on the graph. Another is the clustering coefficient C , obtained by considering the set A_x of neighbors of a vertex x and the fraction C_x of edges between pairs in A_x compared to the total possible: C is the average of C_x . Regular lattices have a big clustering coefficient and short average distance, while random graphs of the Erdős–Renyi type have a small clustering coefficient and short average distance. However, it has been noticed that real networks have a small average distance, but a large clustering coefficient. Thus several examples have been built of graphs exhibiting these features: Watts and Strogatz obtained the first example by starting from an ordered set of n vertices, connecting all vertices with indices closer than a certain distance k and then randomly rewiring the edges one after the other with some probability $\beta \in [0, 1]$; rewiring consists of placing the second end point of an edge uniformly on the remaining vertices not creating loops or double wiring. For $\beta = 0$ the lattice is regular and for $\beta = 1$ it is of Erdős–Renyi type. For $N \gg k \gg \log N$ and intermediate values of β clustering is still close to that of regular lattice and average distance is close to that of random graphs.

Example 8 (SEIR on a Watt–Strogatz small-world network). Starting with the network of Example 1, the network is realized by randomly rewiring with probability δ , using the order 0, a , c , b , the edge which connects the vertex to the successive one and replacing it uniformly where possible. Calculations are quite long and the result is

$$\begin{aligned}
 Prob(R_\infty = \{0\}) &= \int_0^\infty \mu e^{-\mu h} ((1 - p(h)) \left(\frac{1}{2}(1 - \delta)^3 \delta + \frac{10}{12}(1 - \delta)\delta^3 + \frac{3}{8}\delta^4 \right) \\
 &\quad + (1 - p(h))^2 \left((1 - \delta)^4 + 3(1 - \delta)^3 \delta + \frac{61}{12}(1 - \delta)^2 \delta^2 \right. \\
 &\quad \left. + \frac{59}{24}(1 - \delta)\delta^3 + \frac{9}{16}\delta^4 \right) \\
 &\quad + (1 - p(h))^3 \left(\frac{1}{2}(1 - \delta)^3 \delta + \frac{11}{12}(1 - \delta)^2 \delta^2 \right. \\
 &\quad \left. + \frac{17}{24}(1 - \delta)\delta^3 + \frac{1}{16}\delta^4 \right) dh \\
 &= \frac{\mu}{48(\beta + \mu)(2\beta + \mu)(3\beta + \mu)} \\
 &\quad \times (144\beta^2 + 48\beta^2 \delta - 188\beta^2 \delta^2 + 318\beta^2 \delta^3 - 127\beta^2 \delta^4 \\
 &\quad + 192\beta \mu - 44\beta \delta^2 \mu + 94\beta \delta^3 \mu - 35\beta \delta^4 \mu + 48\mu^2).
 \end{aligned}$$

The first models of small-world graphs are regular enough that one can find explicit expressions for the critical values of the parameters at which a large outbreak takes place [67].

How average distance, clustering, or other features affect the value of the epidemic threshold is not very clear. In general, exact results and numerical analysis seem to suggest that clustering reduces the critical point, although it does not have a clear effect on the final size of the epidemic [21]. These conclusions are, however, challenged by several other features, for instance, correlation of in and out degrees: [62] investigates a small-size network and notices that for such a network to have a lower epidemic threshold than other network structures, there needs to be a positive correlation between the number of links to and from nodes. When this correlation is negative the epidemic threshold for small-size networks can be higher [62] concludes that clustering does not necessarily have an influence on the epidemic threshold if connectance is kept constant and that analyses of the influence of the clustering on the epidemic threshold in directed networks can also be spurious if they do not consider simultaneously the effect of the correlation coefficient between in- and out-degree. Finally, it is observed in [60] that clustering by itself does not seem to reduce the epidemic threshold, but that this is the effect of assortativity, the tendency for nodes to contact nodes of similar degree, which is generally associated with clustering.

2.3.3 *Scale-Free Networks*

At the same time, another feature has been considered: the degree distribution of the vertices of the graph decays exponentially in Erdős–Renyi random graphs and in WS but appears to have a power law decay in some realistic network, such as the Internet. Random graphs with this property are called scale-free [24]. A possible construction is by preferential attachment, a procedure in which edges are randomly selected from one vertex with a probability favoring a second vertex with already higher degree [13, 1, 38, 22].

Epidemics has been widely studied in scale-free networks and the surprising feature that simulations and approximations have suggested at first is that the epidemic threshold for a large outbreak is zero [61, 48, 49, 86, 13]. Consequences would be impressive: in a large such network a virus can spread and create a large epidemic no matter how small the individual-to-individual infection probability is.

The issue has not been fully clarified, however. [85] introduces a new scale-free network, which is built recursively by substitutions and studies it asymptotically for a large number of iterations. In such a limit there is a nonzero critical value λ_c for the appearance of a giant component (i.e., for the epidemic outbreak). Another nonzero critical point for a spatial scale-free network is in [72].

Here is an example of preferential attachment with just four individuals even if such a small size induces an excessive oversimplification.

Example 9 (SEIR on a preferential attachment network). Take a random graph $G' = (V', \mathcal{E}')$ with probability δ of each edge in $V' = \{a, b, c\}$; then attach 0 to each individual $i \in V'$ with probability 0 if i is isolated in G' , $1/2$ if i has one neighbor in G' , and 1 if i has two neighbors in G' . From such a random network G an SEIR epidemic with probability $p(h) = 1 - e^{-\beta h}$ of transmission is run starting from 0. Then

$$\begin{aligned}
 \text{Prob}(R_\infty = \{0\}) &= \int_0^\infty \mu e^{-\mu h} \left((1-\delta)^3 + \frac{3}{4}(1-\delta)^2\delta + \frac{3}{8}(1-\delta)\delta^2 \right. \\
 &\quad \left. + (1-p(h)) \left(\frac{3}{2}(1-\delta)^2\delta + \frac{9}{8}(1-\delta)\delta^2 \right) \right. \\
 &\quad \left. + (1-p(h))^2 \left(\frac{3}{4}(1-\delta)^2\delta + \frac{9}{8}(1-\delta)\delta^2 \right) \right. \\
 &\quad \left. + (1-p(h))^3 \left(\frac{3}{8}(1-\delta)\delta^2 + \delta^3 \right) \right) dh \\
 &= \frac{1}{4(\beta + \mu)(2\beta + \mu)(3\beta + \mu)} \\
 &\quad \times (\beta^3(24 - 54\delta + 45\delta^2) + 2\beta^2\mu(22 - 27\delta + 21\delta^2 - 12\delta^3) \\
 &\quad - 3\beta\mu^2(-8 + 4\delta - 5\delta^2 + 5\delta^3) + \mu^3(4 + 3\delta^2 - 2\delta^3)).
 \end{aligned}$$

2.3.4 Spatial Networks and Long-Range Percolation

As they are mostly mass-action models, random graphs, small-world, and scale-free networks do not capture the geographical dispersion of individuals, often in two- or three-dimensional environments, which was a key ingredient of most percolation models. The simplest percolation models, on the other hand, are not scale-free nor have the appropriate clustering coefficient or average distance.

Models which combine dimensionality and nonlocal connections are those named spatial networks: individuals are positioned within a given area (or volume) and two individuals are connected with a probability that depends on their distance. By changing the distribution of individuals or the connection probabilities, it is possible to generate a wide variety of networks, from highly clustered lattices to small-world arrangements to globally connected random networks [35, 71, 45]. They are, however, generally not scale-free, a limitation we address later.

The percolation counterpart of spatial networks is long-range percolation, which is generally defined as a random graph which has $V = \mathbb{Z}^d$ as vertex set and an edge between two vertices at distance r with probability $p_{(r)}$, all edges being selected independently. In general, $p_{(r)}$ only depends on the distance, for instance

$$p(r) = 1 - e^{-\lambda r}, \quad \text{with} \quad \lambda(r) \approx q r^{-s} \quad (2.1)$$

for some $s > 0$ and some q which could be a constant or a slowly varying function of r . In long-range percolation the probability of occurrence of an infinite cluster follows a particular pattern depending on s . In dimension d , if $s < d$ there is an infinite percolation cluster with probability one, while if $s > d$ there is a critical value for q : at small q there is no percolation, but this does occur for large q [19]. Such behavior occurs even in dimension $d = 1$, in which case the ordinary Bernoulli model exhibits no percolation, even if the nontrivial critical point for q appears only for $s \in (1, 2)$ see [74].

Clearly long-range percolation can be considered on other reference networks: an example on a hierarchical lattice is described, for instance, in [53].

Reed–Frost epidemics on the long-range percolation network are studied in [79] and [80]: this gives us the occasion to discuss the role of \mathbf{R}_0 . We have already seen that in spatial graphs the condition $\mathbf{R}_0 = 1$ no longer identifies the onset of a large outbreak. An epidemic model based on a long-range percolation network could be a good place to try an alternative definition of \mathbf{R}_0 . As an attempt, the set B_k of vertices at graphical distance k from the origin is considered in [80], together with the quantities

$$\mathbf{R}_*^{(1)} = \liminf_{k \rightarrow \infty} E(|B_k|)^{1/k}, \quad \mathbf{R}_*^{(2)} = \limsup_{k \rightarrow \infty} E(|B_k|)^{1/k}$$

(using \mathbf{R}_* if the two coincide). It is then shown that for $\lambda(r) \approx r^{-\beta}$ the following occurs: for $\beta < d$ or $\beta = d$ and some other conditions then $\mathbf{R}_* = \infty$; for $\beta = d$ and some other conditions then $1 < \mathbf{R}_*^{(1)} \leq \mathbf{R}_*^{(2)} < \infty$; finally, for $\beta > d$ then $\mathbf{R}_* = 1$ (see also [18]). This could be a first indicator of the epidemic transition to a large outbreak, although it is not able to identify a transition from a small to a large outbreak if $\beta \in (d, 2d)$. The issue is thus still unresolved and for the time being one more easily identifies the onset of a large outbreak by indicating the value of \mathbf{R}_0 at which this occurs.

Example 10 (SEIR epidemics on long-range percolation). The edges $\{0, a\}$ and $\{0, b\}$ are in \mathcal{E} with probability $p_{(1)} = q/1^2 = q$ and $\{0, c\} \in \mathcal{E}$ with probability $p_{(2)} = q/2^2 = q/4$, all independent of one another. With $p(i, j, h) = p(h) = 1 - e^{-\beta h}$ we have

$$\begin{aligned} \text{Prob}(R_\infty = \{0\}) &= \int_0^\infty \mu e^{-\mu h} \sum_{k=0}^2 \binom{2}{k} q^k (1-q)^{2-k} \left(\frac{q}{4} (1-p(h)) \right)^{k+1} \\ &\quad + \left(1 - \frac{q}{4} \right) (1-p(h))^k \Big) dh \\ &= \sum_{k=0}^2 \binom{2}{k} q^k (1-q)^{2-k} \left(\frac{q}{4} \frac{\mu}{\mu + \beta(k+1)} + \left(1 - \frac{q}{4} \right) \frac{\mu}{\mu + \beta k} \right). \end{aligned}$$

As we mentioned, spatial networks and long-range percolation are, however, generally not scale-free: this led Yukich [84] to introduce another class of networks in which a random number is selected for each vertex and then two vertices are connected if their distance is less than both the random numbers. The idea is appealing, but does not seem directly suitable to study epidemics: we shall exploit a similar direction after introducing modularity.

2.3.5 Modular Networks with Communities

It has become clear in the last few years that there is an even more relevant feature of real populations, in particular human ones: the modular or community structure, in which individuals have contacts through different communities to which they belong and, what is more relevant, contacts and thus infectious rates depend on the community [69].

The first step in this direction is to consider, next to usual contacts, a family or household structure in which infections occur at a faster rate [15, 16, 10]. In particular, [11] and [12] study a random network with m households each with n individuals and global random interactions. The spread of infection has then two rates, one for the global and one for the local connections. The epidemic threshold is determined by an extension of the basic reproduction number: suppose there is an initial infected individual 0 and let C_0 be the number of individuals, not in its household, directly infected by 0, T the number of those eventually infected within the household, and C_1 those directly infected by any individual different from 0 in its household. Then the onset of a large outbreak depends on whether $\mathbf{R}_* > 1$ or $\mathbf{R}_* < 1$, with $\mathbf{R}_* = E(C_0) + E(T)E(C_1)$. This is to be contrasted with the usual \mathbf{R}_0 which is between $E(C_0)$ and $E(C_0) + E(T)$.

Example 11 (SEIR epidemics with households). Assume that $\{0, a, b\}$ are in the same household and their connecting edges are automatically in \mathcal{E} . The individual in c is outside the household and there is some probability $p_{(c)}$ that $\{0, c\} \in \mathcal{E}$. The infection rate β_H within the household is higher than the infection rate $\beta_{(c)}$ between 0 and c

$$\begin{aligned}
 Prob(R_\infty = \{0\}) &= \int_0^\infty \mu e^{-\mu h} ((1 - p_{(c)})(1 - p(0, a, h))(1 - p(0, b, h)) \\
 &\quad + p_{(c)}(1 - p(0, a, h))(1 - p(0, b, h))(1 - p(0, c, h)) dh \\
 &= \int_0^\infty \mu e^{-\mu h} \left((1 - p_{(c)})e^{-2\beta_H} + p_{(c)}e^{-(2\beta_H + \beta_{(c)})} \right) dh \\
 &= (1 - p_{(c)}) \frac{\mu}{\mu + 2\beta_H} + p_{(c)} \frac{\mu}{\mu + 2\beta_H + \beta_{(c)}}
 \end{aligned}$$

In another direction, several papers have investigated the modular structure of established networks to evaluate the effects on epidemics: [66] presents a class of new divisive algorithms for the discovery of community structure in networks and gives references to other existing methods; [65] shows that the modularity can be expressed in terms of the eigenvectors of a characteristic matrix for the network, named modularity matrix, and by this expression generates a spectral algorithm for community detection (see also [31] and [70]).

Modular effects on epidemics are directly studied in [58] which reports indications, by exact formulas and simulations, that the epidemic threshold decreases with community structures. [30] shows that internal links within the communities are less responsible of the spreading of a disease with respect to external ones. In a slightly different model, dynamical behaviors of epidemics on scale-free networks with community structure is considered in [78] where an SIR model is studied, with communities and movement of individuals away from infected communities.

An interesting network with modular as well as hierarchical structure is proposed in [68]. It starts from a modular network consisting of m modules, each containing n nodes. The idea is to take a rich structure within each module, and then analyze its modification under the hierarchical construction. Within each module the connectivity (i.e., the probability of a link between any pair of nodes) is taken to be some ρ_1 , while the connectivity between modules is $\rho_2 = \rho \rho_1$, with $\rho \in [0, 1]$. The hierarchy is introduced by adding another set of m modules (each having n nodes) with the same ρ_1 and ρ_2 . The nodes belonging to these two different sets of modules are then connected, but with a probability $\rho_3 = \rho^2 \rho_1$. The resulting network has $2nm$ nodes and $l = 2$ hierarchical levels. To increase the number of hierarchical levels to $l = 3$, a similar network is added with $2nm$ nodes to the existing network and, as above, links between these two networks occur with a probability $\rho_4 = \rho^3 \rho_1$. Thus, to get a network with $l = h$ hierarchical levels, the above procedure is repeated $h - 1$ times. The final network contains $M = 2(h - 1)m$ number of modules. No study of epidemics has been carried out on the network, and, in any case, the random network has no clear spatial features.

2.4 A New Class of Epidemic Models and Their Percolation Analysis

2.4.1 *Epidemics on Modular Spatial Hierarchical Scale-Free Networks*

Gathering several of the previous considerations we propose here a new class of random networks called SHEM which is likely to capture important features of epidemics in human communities: networks in the class have spatial dimensionality, modularity, are built with a hierarchical structure, and can be made scale-free by a suitable choice of the parameters. The class is likely to be flexible enough to

model several different situations but is also likely to be regular enough that the analysis of most examples will not be too difficult. Here is a general description of the class, followed by one detailed example.

To construct the SHEM random network, let the vertex set of the graph be $V = \mathbb{Z}^d$ or any regular finite or infinite lattice. The set of edges \mathcal{E} contains all the nearest neighbor edges of V plus a random set of edges defined as follows. Consider a probability distribution ν on $H = \{0, 1\}^{\mathbb{N}}$, and for each $i \in V$ an independent realization $\eta(i) \in H$. Next, for each $k \in \{0, 1, \dots\}$ partition V into blocks $B_r(k)$ of level k ; although easier to picture, blocks are not necessarily hypercubes. Then connect two vertices $i, j \in V$ if there exists a level k and a block $B_r(k)$ such that $i, j \in B_r(k)$ and $\eta_k(i) = \eta_k(j) = 1$; multiple edges can be removed or not, according to the desired model. Clearly, the number of levels is finite in any realistic example, with the largest block being the entire population, but can be taken infinite for mathematical convenience.

Vertices $i \in V$ represent individuals, and blocks $B_r(k)$ represent communities. For instance, blocks at level 1 may represent families, at level 2 classes or workplaces, at level 3 schools or companies. At higher levels, blocks may represent special communities which are accessed by only few individuals selected out of a large potential set of members: sports teams, political bodies, company boards, etc. By appropriate selection of the forms of the blocks and of the distribution ν one can reproduce the community distributions of human populations of any size. Edges in the graph represent the closeness of individuals and the opportunity of infectious contacts in the case of epidemic. Nearest-neighbor edges in \mathbb{Z}^d represent the pairs of individuals which are extremely close by, whether or not they are in the same family; note that in $d = 2$ the number of such close relations is 4. Edges within communities, which connect all members of the community, represent the opportunity of disease transmission within the community itself.

Once the random network $G = (V, \mathcal{E})$ is constructed one can study or simulate a diffusion of an infectious disease on G , along the edges both between nearest neighbors and within the communities. Once an individual is infected, transmission to neighbors in the random network takes place at some rate, which can be different for different pairs: it actually makes sense to have community-dependent rates of transmission to the susceptible individual belonging to the same community. A natural assumption is that transmission between a pair of individuals is attempted independently in each shared community, increasing the transmission rate if many communities are shared. It makes sense, also, that the rate at which transmission takes place decreases with the size of the community, as direct contacts become more rare. All of these rates may even depend on infectivity and susceptibility.

To simplify things in this presentation, we restrict ourselves now to a specific form of the distribution ν : we fix a nonnegative integer distribution P with probabilities p_k , $k \in \{0, 1, \dots\}$, and for each vertex $i \in V$ consider an independent realization X_i with distribution P ; $\eta_k(i) = 1$ if $X_i \geq k$ and else $\eta_k(i) = 0$. This creates a strong and unnatural dependence between the different communities containing a given individual, but it makes the mathematical analysis much easier, while still capturing some features of human communities.

Example 12 (SEIR epidemics in the hierarchical random network with communities). Assume that the transmission probabilities $p(0, i, h) = p(h)$ are independent of i but are higher within the household, and that X_i satisfy $P(X_i = 0) = 2/3 = 1 - P(X_i \geq 1)$. a and b are also the nearest neighbors of 0. Then

$$\begin{aligned} \text{Prob}(R_\infty = \{0\}) &= \frac{2}{3} \int_0^\infty \mu e^{-\mu h} ((1 - p_H(h))^2 dh \\ &\quad + \frac{1}{3} \sum_{r=0}^3 \binom{3}{r} \int_0^\infty \mu e^{-\mu h} \left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{3-r} ((1 - p_H(h))^2 ((1 - p(h))^r dh \\ &= \frac{2}{3} \frac{\mu}{\mu + 2\beta_H} + \frac{1}{3} \sum_{r=0}^3 3r \left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{3-r} \frac{\mu}{\mu + 2\beta_H + r\beta} \end{aligned}$$

The random network proposed in this section is reminiscent of the one introduced in [84], described above at the end of Sect. 2.3.4, which is obtained as a limit case (see below), but captures more features of the spread of infectious diseases. Similar ideas without spatial dimensions appear in [25] and [75]. Note that the random values X_i 's assigned to each individual to construct the graph play a different role from that assigned to infectivity and susceptibility in [59]. The key distinguishing element is that in the proposed model the variable X_i 's are used both to generate the random network through the blocks and to determine the intensity of the interactions via the number of shared blocks.

2.4.2 Percolation Analysis of a Specific Case

Here is a specific example within the above class, which is further studied in detail in [29]. Consider dimension $d = 2$, $p_k = (\alpha - 1)\alpha^{-(k+1)}$, $k = 0, 1, \dots$, and nested blocks $\{B_{r,s}(k)\}$, $r, s \in \mathbb{Z}$, of size 2^k of the form

$$B_{r,s}(k) = \{(i_1, i_2) \in \mathbb{Z}^2 \mid r2^k \leq i_1 \leq (r+1)2^k - 1, s2^k \leq i_2 \leq (s+1)2^k - 1\}.$$

A random network G is then created as described above, by removing multiple edges and adding all nearest neighbor edges in \mathbb{Z}^2 . The rigidity of the blocks does not allow to reproduce the community distribution for the low level blocks, such as the household distribution <http://www.statistics.gov.uk/census/>. But it correctly reproduces asymptotic features, such as the distribution of community sizes: this is in general polynomial, with a fraction of communities of size m of the order of c/m^5 [56]. In fact, the average number of members of a community at level k is $\frac{2^k}{\alpha^k}$, and thus a community has size m at level $k(m) = \frac{\log m}{\log 4 - \log \alpha}$. If we consider a

population of $2^{2\bar{k}}$, there are $\frac{2^{2\bar{k}}}{2^{2k(m)}} = \frac{2^{2\bar{k}}}{m^{\frac{2\log 2}{\log 4 - \log \alpha}}} \approx c/m^{4.8}$ communities of size m . The average vertex degree \bar{d} of the origin satisfies

$$\bar{d} = 4 + \sum_{k=2}^{\infty} \frac{1}{\alpha^{2k}} \left(2^{2k} - 2^{2(k-1)} \right) = 4 + \frac{3}{\alpha^2 - 4}$$

for $\alpha = 3$ we have $\bar{d} \approx 4.6$. For $\gamma - 1 = \frac{\log_2 \alpha}{2 - \log_2 \alpha}$ the degree distribution satisfies $P(d \geq k) \approx ck^{-(\gamma-1)}$ for some constant $c > 0$, hence $P(d = k) \approx ck^{-\gamma}$ and the random network is scale-free for $2 \leq \alpha \leq 2^{\frac{4}{3}}$. For $\alpha = 3$ we have $P(d = k) \approx c/k^{2.81}$ [29].

To perform a simple analysis of an epidemics on the random network G , we take a Reed–Frost model, i.e. the SEIR epidemic model with constant infectious periods and transmission rates not depending on the individual. This amounts to a Bernoulli percolation process on G . Transmission probabilities, however, depend on the community shared by two individuals. There is a transmission probability p between nearest neighbors, so that at the level of the nearest neighbor graph a standard Bernoulli percolation takes place. Then within each community $B_{r,s}(k)$ we assume that the probability of transmission is $p\rho^k$, $\rho \in [0, 1]$, with transmissions independent for the various individuals and for the different communities even between the same individuals. Therefore, once one of i or j is infected and the other is susceptible, then the probability of transmission is $p(i, j, X_i, X_j) = 1 - \prod_{k=0}^{\infty} (1 - p\rho^k \mathbb{I}_{\{\{i,j\}|\exists r,s:X_i,X_j \geq k \text{ and } i,j \in B_{r,s}(k)\}})$.

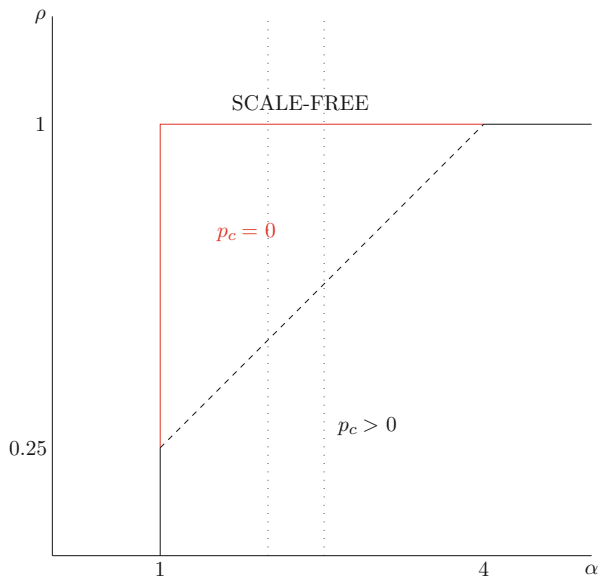
In the limit $\alpha \rightarrow 1$, $p_k \rightarrow 0$ and $P(X \geq k) \rightarrow 1$ for all k : in practice $X_i \equiv \infty$, each individual belongs to all communities and if $m \approx \log_2 r(i, j)$, where $r(i, j)$ is the distance between i and j , is the linear size of the first block to which i and j belong

$$\begin{aligned} p(i, j) &= 1 - \prod_{k=m}^{\infty} (1 - p\rho^k) \\ &\approx 1 - e^{-\sum_{k=m}^{\infty} p\rho^k} \\ &\approx c p \rho^m \approx c p \rho^{\log_2 r(i, j)} \approx c p r(i, j)^{\log \rho / \log 2} \end{aligned}$$

for some constant $c > 0$. Therefore, the model is close to the long-range percolation model on \mathbb{Z}^2 with $s = -\log \rho / \log 2$ and $q = cp$. Percolation, and hence a large outbreak in the Reed–Frost epidemics, occurs at all values of p if $s = -\log \rho / \log 2 < 2$, i.e. $\rho > 1/4$, while there is a nontrivial transition in p between small and large outbreak if $\rho < 1/2^d = 1/4$: the model is studied in detail in [53, 34, 7]. It is now interesting to investigate if this behavior occurs also in the scale-free parameter region.

It turns out that for $\alpha > 4$ the nearest neighbor transmissions take over and thus the model is basically equivalent to the short-range Bernoulli percolation discussed at the end of Sect. 2.2.4 [29]. In the limit $\alpha \rightarrow \infty$ only nearest neighbor edges

Fig. 2.6 $\alpha - \rho$ phase diagram of the hierarchical modular spatial random network in dimension 2



remain, so that we obtain the standard Bernoulli bond percolation in \mathbb{Z}^2 , with the exact critical point $p_c = 1/2$ (see Sect. 2.2.4). Also for $\rho = 0$ one retrieves the standard Bernoulli bond percolation in \mathbb{Z}^2 , as no transmission can occur at any higher level. For $\rho = 1$ the model has some independent interest: two vertices i and j are connected if and only if $X_i, X_j \geq k(i, j)$ where $k(i, j)$ is the size of the smallest block containing them. The model resembles that of Yukich (see end of Sect. 2.3.4), but with some differences due to the block structure. Thus the model presented here interpolates between long-range and short-range percolation, and in part of the interpolating region it is scale-free.

For $1 < \alpha < 4$ the behavior is similar to that of long-range percolation: if $\rho < \alpha/4$ there is a nontrivial transition from small to large outbreak, while for $\rho > \alpha/4$ the large outbreak occurs for all values of p : a detailed proof involves several technicalities [29].

As this behavior takes place also in the scale-free region it clarifies the claim that in scale-free networks there could be no small outbreak at all transmission rates: this is indeed the case but only if the decrease in transmission rates in large communities is too slow. These calculations suggest various lines of intervention, such as reducing transmission rates in large communities or vaccinating their more prominent members (who would also belong to even larger communities).

These results are so far only at qualitative level but once more appropriate structures of low level blocks and more realistic transmission mechanisms are introduced, they can be used for quantitative evaluations of epidemics (Fig. 2.6).

2.5 Conclusions

We have seen that there is a rich and nontrivial interplay between epidemic and percolation models; such interplay takes advantage of suitable transformations which eliminate time-dependent random variables and may provide exact or approximate relations. On the other hand, it is desirable to introduce new individually based random variables to describe features like the susceptibility of individuals or the structure of the network, and percolation can often encompass such new sources of random behavior.

In particular, we have proposed a new family of random networks which incorporate spatial features, modularity and scale-free characteristics using individual-based random variables. Various types of percolation appear as limiting cases of the new model. The study of the additional features has been carried out, however, only for one specific example, and it would be interesting to study the general stochastic SEIR model on such random networks and analyze the corresponding percolation models. The new model has also been presented in rather general terms, and one needs to develop statistical procedures like model selection, network determination, parameter estimation, simulation and prediction.

The above considerations support the view that the interplay between percolation and epidemics is still at an early stage.

References

1. Albert R, Jeong H, Barabasi A-L (1999) Diameter of the World Wide Web. *Nature* 401:130
2. Allen LJS (2008) An introduction to stochastic epidemic models. Summer School on Mathematical Modeling of Infectious Diseases, University of Alberta Lecture Notes
3. Allen LJS, Fienberg SE, Holland PW (2008) An introduction to stochastic epidemic models. Springer, Berlin
4. Anderson RM, May RM (1991) Infectious diseases of humans. Oxford University Press, Oxford
5. Andersson H (1999) Epidemic models and social networks. *Math Scientist* 24:128
6. Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. Springer Lecture Notes in Statistics. Springer, New York
7. Athreya SR, Swart JM (2011) Survival of contact processes on the hierarchical group. Preprint. arXiv:0808.3732v3
8. Bailey NTJ (1975) The mathematical theory of infectious diseases and its applications, 2nd edn. Griffin, London
9. Balister PN, Bollobas B (2005) Continuum percolation in the square and the disk. *Random Struct Algor* 26:392–403
10. Ball FG, Mollison D, Scalia-Tomba G (1997) Epidemics with two levels of mixing. *Ann Appl Probab* 7(1):46–89
11. Ball F, Sirl D, Trapman P (2009) Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv Appl Probab* 41:765–796
12. Ball F, Sirl D, Trapman P (2010) Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math Biosci* 224(2):53–73

13. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
14. Barbour AD (1975) The duration of the closed stochastic epidemic. *Biometrika* 62:477–482
15. Bartoszyński R (1972/73) On a certain model of an epidemic. *Zastos Mat* 13:139–151
16. Becker NG, Dietz K (1995) The effect of household distribution on transmission and control of highly infectious diseases. *Math Biosci* 127:207–219
17. Beffara V, Sidoravicius V (2006) Percolation. *Encyclopedia of mathematical physics*, vol 4. Elsevier, Amsterdam, pp 2120–2126
18. Benjamini I, Berger N (2001) The diameter of long-range percolation clusters on finite cycles. *Random structures and algorithms* 19:102–111
19. Berger N (2002) Transience, recurrence and critical behavior for long-range percolation. *Commun Math Phys* 226:531–558
20. Britton T (2005) *Stochastic epidemic models: a survey*. Cambridge University Press, New York.
21. Britton T, Deijfen M, Lagers AN, Lindholm M (2008) Epidemics on random graphs with tunable clustering. *J Appl Probab* 45(3):743–756
22. Britton T, Janson S, Martin-Löf A (2007) Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Adv in Appl Probab* 39(4):922–948
23. Broadbent S, Hammersley J (1957) Percolation processes I. Crystals and mazes. *Proc Cambridge Philos* 53:629–641
24. Caldarelli G (2007) *Scale-Free networks complex webs in nature and technology*. Oxford University Press, Oxford
25. Caldarelli G, Capocci A, De Los Rios P, Munoz MA (2002) Scale-Free networks from varying vertex intrinsic fitness. *Phys Rev Lett* 89:258702
26. Camia F, Newman CM (2006) Two-dimensional critical percolation: the full scaling limit. *Comm Math Phys* 268(1):1–38
27. Cardy J (2008) Conformal field theory and statistical mechanics. *Exact methods in low-dimensional, statistical physics and quantum computing*. Les Houches Summer School Lectures.
28. Cardy JL, Grassberger P (1985) Epidemic models and percolation. *J Phys A-Math Gen* 18:L267–L271
29. Cecconi L, Gandolfi A (2011) SIR epidemics on a scale-free spatial nested modular network. *arXiv:1107.1532*
30. Chu X, Guan J, Zhang Z, Zhou S (2009) Epidemic spreading in weighted scale-free networks with community structure. *J Stat Mech-Theory E* 2009(07):P07043
31. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E* 70(6):066111
32. Daley DJ, Gani J (2001) *Epidemic modelling: an introduction*. Cambridge University Press, Cambridge, UK
33. Davis S, Trapman P, Leirs H, Begon M, Heesterbeek JAP (2008) The abundance threshold for plague as a critical percolation phenomenon. *Nature* 454:634–637
34. Dawson D, Gorostiza L (2011) Percolation in an ultrametric space. Preprint. *arXiv:1006.4400v2*
35. Eames KTD, Keeling MJ (2002) Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc Natl Acad Sci USA* 99:13330–13335
36. Erdős P, Rényi A (1959) On random graphs, I. *Publicationes Mathematicae (Debrecen)* 6:290–297
37. Erdős P, Rényi A (1960) The evolution of random graphs. *Magyar Tud Akad Mat Kut Int Klizleményei* 5:17–61
38. Eriksen KA, Hornquist M (2001) Scale-free growing networks imply linear preferential attachment. *Phys Rev E* 65(1):017102
39. Gandolfi A, Keane M, De Valk V (1989) Extremal two-correlations of two-valued stationary one-dependent processes. *J Probab Theory Rel* 80:475–480

40. Garett O, Marchand R (2004) Asymptotic shape for the chemical distance and first-passage percolation in random environment. *ESAIM: Probab Statist* 8:169–199
41. Grassberger P (1983) On the critical behaviour of the general epidemic process and dynamical percolation. *Math Biosci* 63:157–172
42. Grimmett GR (1999) Percolation. vol. 321 of *Grundlehren der Mathematischen Wissenschaften*, 2nd edn. Springer, Berlin
43. Gutfraind A (2010) Monotonic and non-monotonic epidemiological models on networks. Preprint. arXiv:1005.3470v2
44. Hethcote HW (2000) The mathematics of infectious diseases. *J Soc Ind Appl Math* 42:599–653
45. Keeling MJ (2005) Implications of network structure for epidemic dynamics. *Theor Popul Biol* 67:1–8
46. Kenah E, Miller JC (2011) Epidemic percolation networks, epidemic outcomes, and interventions. *Interdiscip Perspect Infect Dis* 2011:1–13
47. Kenah E, Robins JM (2007) Second look at the spread of epidemics on networks. *Phys Rev E* 76(3):036113
48. Kephart JO, Sorkin GB, Chess DM et al (1997) Fighting computer viruses. *Sci Am* 277:56–61
49. Kephart JO, White SR, Chess DM (1993) Computers and epidemiology. *IEEE Spectr* 30:20–26
50. Kermack W, McKendrick A (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc London A* 115:700–721
51. Kesten H (1980) The critical probability of bond percolation on the square lattice equals $1/2$. *Comm Math Phys* 74:41–59
52. Kesten H (1982) Percolation theory for mathematicians. *Progress in Probability and Statistics*, vol. 2, Birkhauser, Boston
53. Koval V, Meester R, Trapman P (2011) Long-range percolation on the hierarchical lattice. Preprint. arXiv:1004.1251v1
54. Kuulasmaa K (1982) The spatial general epidemic and locally dependent random graphs. *Appl Probab* 19:745–758
55. Kuulasmaa K, Zachary S (1984) On spatial general epidemics and bond percolation processes. *J Appl Prob* 21(4):911–914
56. Lancichinetti A, Kivela M, Saramaki J, Fortunato S (2010) Characterizing the community structure of complex networks. *PLoS One* 5:e11976
57. Lefèvre C, Picard P (1990) A non-standard family of polynomials and the final size distribution of Reed-Frost epidemic processes. *Adv Appl Prob* 22:25–48
58. Liu ZH, Hu BB (2005) Epidemic spreading in community networks. *Europhys Lett* 72:315
59. Meester R, Trapman P (2010) Bounding basic characteristics of spatial epidemics with a new percolation model. Preprint.
60. Miller J (2007) Predicting the size and probability of epidemics in populations with heterogeneous infectiousness and susceptibility. *Phys Rev E* 76 010101(R)
61. Moreno Y, Gómez JB, Pacheco AF (2003) Epidemic incidence in correlated complex networks. *Phys Rev E* 68(3):035103
62. Moslonka-Lefebvre M, Pautassoc M, Jeger MJ (2009) Disease spread in small-size directed networks: epidemic threshold, correlation between links to and from nodes, and clustering. *J Theor Biol* 260(3):402–411
63. Neal P (2003) SIR epidemics on a bernoulli random graph. *J Appl Probab* 40(3):779–782
64. Neal P, Martin-Löf A (1986) Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J Appl Probab* 23(2):265–282
65. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577
66. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69(2):026113
67. Newman MEJ, Watts DJ (1999) Scaling and percolation in the small-world network model. *Phys Rev E* 60:7332–7342

68. Pan RK, Sinha S (2008) Modular networks with hierarchical organization: the dynamical implications of complex structure. *Pramana: J Phys* 71(2008):331–340
69. Pellis L, Ferguson NM, Fraser C (2011) Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *J Math Biol* 63(4):691–734
70. Radicchi F, Castellano C, Ceconi F, Loreto V, Parisi D (2004) Defining and identifying communities in networks. *Proc Natl Acad Sci USA* 101(9):2658–2663
71. Read JM, Keeling MJ (2003) Disease evolution on networks: the role of contact structure. *Proc R Soc B* 270:699–708
72. Sander LM, Warren CP, Sokolov IM (2003) Epidemics, disorder, and percolation. *Physica A* 325(1):1–8
73. Sander LM, Warren CP, Sokolov IM, Simon C, Koopman J (2002) Percolation on heterogeneous networks as a model for epidemics. *Math Biosci* 180:293–305
74. Schulman LS (1983) Long range percolation in one dimension. *J Phys A Lett* 16:L639–L641
75. Servedio VDP, Buttà P, Caldarelli G (2004) Vertex intrinsic fitness: how to produce arbitrary scale-free networks. *Phys Rev E* 70(5):056126
76. Smirnov S (2005) Critical percolation and conformal invariance. In: XIVth International Congress on Mathematical Physics. World Scientific Publishing, Hackensack, pp 99–112
77. Stauffer D, Aharony A (1994) Introduction to percolation theory, 2nd edn. Taylor and Francis, London
78. Suna HJ, Gao ZY (2007) *Physica A: Statistical Mechanics and its Applications* 381:491–496
79. Tan Z-J, Zou X-W, Jin Z-Z (2000) Percolation with long-range correlations for epidemic spreading. *Phys Rev E* 62:8409–8412
80. Trapman P (2010) The growth of the infinite long-range percolation cluster. *Ann Probab* 38(4):1583–1608
81. Van den Berg J, Grimmett GR, Schinazi RB (1998) Dependent random graphs and spatial epidemics. *Ann Appl Probab* 8(2):317–336
82. Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. *Nature* 393(6684):440–442
83. Werner W (2004) Random planar curves and Schramm-Loewner evolutions. In: *Lectures on Probability Theory and Statistics. Lecture Notes in Mathematics*, vol 1840. Springer, Heidelberg, pp 107–195
84. Yukich, JE (2006) Ultra-small scale-free geometric networks. *J Appl Probab* 43:665–677
85. Zhang Z, Zhou S, Zou T, Chen L, Guan J (2009) Different thresholds of bond percolation in scale-free networks with identical degree sequence. *Phys Rev E* 79(3):031110
86. Zhou T, Fu ZQ, Wang BH (2006) Epidemic dynamics on complex networks. *Prog Nat Sci* 16(5):452–457

Dynamic Models of Infectious Diseases

Volume 2: Non Vector-Borne Diseases

Rao, V.S.H.; Durvasula, R. (Eds.)

2013, XII, 259 p. 69 illus., 42 illus. in color., Hardcover

ISBN: 978-1-4614-9223-8