Epidemic processes in complex networks

Romualdo Pastor-Satorras*

Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

Claudio Castellano

Istituto dei Sistemi Complessi (ISC-CNR), via dei Taurini 19, I-00185 Roma, Italy and Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy

Piet Van Mieghem

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Alessandro Vespignani

Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston Massachusetts 02115, USA and Institute for Scientific Interchange Foundation, Turin 10133, Italy

(published 31 August 2015)

In recent years the research community has accumulated overwhelming evidence for the emergence of complex and heterogeneous connectivity patterns in a wide range of biological and sociotechnical systems. The complex properties of real-world networks have a profound impact on the behavior of equilibrium and nonequilibrium phenomena occurring in various systems, and the study of epidemic spreading is central to our understanding of the unfolding of dynamical processes in complex networks. The theoretical analysis of epidemic spreading in heterogeneous networks requires the development of novel analytical frameworks, and it has produced results of conceptual and practical relevance. A coherent and comprehensive review of the vast research activity concerning epidemic processes is presented, detailing the successful theoretical approaches as well as making their limits and assumptions clear. Physicists, mathematicians, epidemiologists, computer, and social scientists share a common interest in studying epidemic spreading and rely on similar models for the description of the diffusion of pathogens, knowledge, and innovation. For this reason, while focusing on the main results and the paradigmatic models in infectious disease modeling, the major results concerning generalized social contagion processes are also presented. Finally, the research activity at the forefront in the study of epidemic spreading in coevolving, coupled, and time-varying networks is reported.

DOI: 10.1103/RevModPhys.87.925 PACS numbers: 05.10.−a, 64.60.aq, 89.75.−k

CONTENTS

*romualdo.pastor@upc.edu

I. INTRODUCTION

Since the first mathematical approach to the spread of a disease by Daniel Bernoulli (1760), epidemic models lie at the core of our understanding about infectious diseases. As experimenting *in vivo* epidemics is not a viable option, modeling approaches have been the main resort to compare and test theories, as well as to gauge uncertainties in intervention strategies. The acclaimed work of Kermack and McKendrick (1927), defining the modern mathematical modeling of infectious diseases, has evolved through the years in an impressive body of work, whose culmination is well represented by the monumental summary of Anderson and May (1992). At the same time, the epidemic modeling metaphor has been introduced to describe a wide array of different phenomena. The spread of information, cultural norms, and social behavior can be conceptually modeled as a contagion process. How blackouts spread on a nationwide scale or how efficiently memes can spread on social networks are all phenomena whose mathematical description relies on models akin to classic epidemic models (Vespignani, 2012). Although the basic mechanisms of each phenomenon are different, their effective mathematical description often defines similar constitutive equations and dynamical behaviors framed in the general theory of reaction-diffusion processes (van Kampen, 1981). It is not surprising then that epidemic modeling is a research field that crosses different disciplines and has developed a wide variety of approaches ranging from simple explanatory models to elaborate stochastic methods and rigorous results (Keeling and Rohani, 2007).

In recent years we witnessed a second golden age in epidemic modeling. Indeed, the real-world accuracy of the models used in epidemiology has been considerably improved by the integration of large-scale data sets and the explicit simulation of entire populations down to the scale of single individuals (Eubank *et al.*, 2004; Ferguson *et al.*, 2005; Longini *et al.*, 2005; Halloran *et al.*, 2008; Balcan, Colizza *et al.*, 2009; Chao *et al.*, 2010; Merler *et al.*, 2011). Mathematical models have evolved into microsimulation models that can be computationally implemented by keeping track of billions of individuals. These models have gained importance in the public-health domain, especially in infectious disease epidemiology, by providing quantitative analyses in support of policy-making processes. Many researchers are advocating the use of these models as real-time, predictive tools (Nishiura, 2011; Tizzoni *et al.*, 2012; Nsoesie *et al.*, 2014). Furthermore, these models offer a number of interesting and unexpected behaviors, whose theoretical understanding represents a new challenge, and have stimulated an intense research activity. In particular, modeling approaches have expanded into schemes that explicitly include spatial structures, individual heterogeneity, and the multiple time scales at play during the evolution of an epidemic (Riley, 2007).

At the core of all data-driven modeling approaches lies the structure of human interactions, mobility, and contact patterns that finds its best representation in the form of networks (Butts, 2009; Vespignani, 2009, 2012; Jackson, 2010; Newman, 2010). For a long time, detailed data on those networks were simply unavailable. The new era of the social web and the data deluge is, however, lifting the limits scientists have been struggling with for a long time. The pervasive use of mobile and wifi technologies in our daily life is changing the way we can measure human interactions and mobility network patterns for millions of individuals at once. Sensors and tags are able to produce data at the microscale of one-to-one interactions. Proxy data derived from the digital traces that individuals leave in their daily activities (microblogging messages, recommendation systems, consumer ratings) allow the measurement of a multitude of social networks relevant to the spreading of information, opinions, habits, etc.

Although networks have long been acknowledged as a key ingredient of epidemic modeling, the recent abundance of data is changing our understanding of a wide range of phenomena and calls for a detailed theoretical understanding of the interplay between epidemic processes and networks. A large body of work has shown that most real-world networks exhibit dynamic self-organization and are statistically heterogeneous—typical hallmarks of complex systems (Albert and Barabási, 2002; Dorogovtsev and Mendes, 2002, 2003; Newman, 2003b, 2010; Boccaletti *et al.*, 2006; Caldarelli, 2007; Costa *et al.*, 2007; Cohen and Havlin, 2010; Baronchelli *et al.*, 2013). Real-world networks of relevance for epidemic spreading are different from regular lattices. Networks are hierarchically organized with a few nodes that may act as hubs and where the vast majority of nodes have few

interactions. Both social and infrastructure networks are organized in communities of tightly interconnected nodes. Although randomness in the connection process of nodes is always present, organizing principles and correlations in the connectivity patterns define network structures that are deeply affecting the evolution and behavior of epidemic and contagion processes. Furthermore, network's complex features often find their signature in statistical distributions which are generally heavy tailed, skewed, and varying over several orders of magnitude.

The evidence of large-scale fluctuations, clustering, and communities characterizing the connectivity patterns of real-world systems has prompted the need for mathematical approaches capable of dealing with the inherent complexity of networks. Unfortunately, the general solution, using, e.g., the master equation of the system, is hardly achievable even for very simple dynamical processes. For this reason, an intense research activity focused on the mathematical and computational modeling of dynamical processes on networks has started across different disciplines (Dorogovtsev, Goltsev, and Mendes, 2008). The study of network evolution and the emergence of macrolevel collective behavior in complex systems follows a conceptual route essentially similar to the statistical physics approach to nonequilibrium phase transitions (Henkel, Hinrichsen, and Lübeck, 2008). Hence, statistical physics has been leading the way to the revamped interest in the study of contagion processes, and more generally dynamical processes in complex networks. Indeed in the last ten years, an impressive amount of methods and approaches ranging from mean-field theories to rigorous results has provided new quantitative insights in the dynamics of contagion processes in complex networks (Keeling and Eames, 2005; Danon *et al.*, 2011).

However, as is often the case in research areas pursued by different scientific communities, relevant results are scattered across domains and published in journals and conference proceedings with completely different readership. In some cases, relevant advances have been derived independently by using different jargons as well as different assumptions and methodologies. This fragmented landscape does not advance the field and is, in many cases, leading to the compartmentalization and duplication of the research effort. We believe that a review is timely to contextualize and relate the recent results on epidemic modeling in complex networks. Although infectious diseases will be at the center stage of our presentation, social contagion phenomena and network dynamics itself are discussed, offering a general mathematical framework for all social and information contagion processes that can be cast in the epidemic metaphor. The final goal is to provide a coherent presentation of our understanding of epidemic processes in populations that can be modeled as complex networks.

After a review of the fundamental results in classical epidemic modeling and the characterization of complex networks, we discuss the different methodologies developed in recent years to understand the dynamic of contagion processes in the case of heterogeneous connectivity patterns. In particular, in Sec. IV we specifically spell out the assumptions inherent to each methodology and the range of applicability of each approach. In Sec. V those theoretical approaches are applied to classic epidemic models such as the susceptible-infected-susceptible (SIS) and susceptible-infected-recovered (SIR) models. In those sections particular care is devoted to shed light on the role of the interplay of the time scales of the epidemic process and of the network dynamics and on the appropriateness of different modeling approximations. In Secs. VI and VII we focus on various approaches to the mitigation and containment of epidemic processes and on the analysis of several variations of the basic epidemic models, aiming at a more realistic description of contagion processes and contact patterns. In Sec. VIII we provide a summary of recent results concerning time-varying networks. Although this is an area that is rapidly advancing due to both theoretical and data gathering efforts, we report on results that are expected to become foundational. In Sec. IX we discuss the generalization of epidemic processes in complex, multispecies reaction-diffusion processes, an area relevant in the analysis of epidemics in structured populations. Finally, in Sec. X, we review the generalization of epidemic modeling of social contagion phenomena. The number of specific models for social contagion is extensive and we therefore confine ourselves to the most relevant to highlight differences and novel dynamical behaviors in the evolution of the epidemic process. We conclude with an outlook on the field and the challenges lying ahead of us.

The upsurge of interest in epidemic modeling in complex networks has led to an enormous body of work: a query on the Thompson Web of Science database with the keywords "epidemic" and "networks" returns more than 3600 papers in just the last 15 years. A review of all these papers is unfortunately hardly feasible. Therefore, we have concentrated our attention on to what we believe are the most influential papers. In providing a unified framework and notation for the various approaches, we aim at fostering synergies across application domains and provide a common knowledge platform for future efforts in this exciting research area.

II. THE MATHEMATICAL APPROACH TO EPIDEMIC SPREADING

A. Classical models of epidemic spreading

In more than 200 years of its history, the mathematical modeling of epidemic spreading has evolved into a research area that spans across several fields of mathematical biology as well as other disciplines and is treated in classic books such as those by Anderson and May (1992), Andersson and Britton (2000), Diekmann and Heesterbeek (2000), Keeling and Rohani (2007), Brauer and Castillo-Chavez (2010), and Diekmann, Heesterbeek, and Britton (2012). Here we merely set the notation and present some of the basic elements and approximations generally used in the modeling of epidemic phenomena in order to provide the necessary conceptual toolbox needed in the following sections.

Epidemic models generally assume that the population can be divided into different classes or compartments depending on the stage of the disease (Anderson and May, 1992; Diekmann and Heesterbeek, 2000; Keeling and Rohani, 2007), such as susceptibles (denoted by S, those who can contract the infection), infectious (I, those who contracted the

infection and are contagious), and recovered or removed[1] (R, those who are removed from the propagation process, either because they have recovered from the disease or because they have died). Additional compartments can be used to signal other possible states of individuals with respect to the disease, for instance immune individuals. This framework can be extended to take into account vectors, such as mosquitoes for malaria, for diseases propagating through contact with an external carrier. Epidemic modeling describes the dynamical evolution of the contagion process within a population. In order to understand the evolution of the number of infected individuals in the population as a function of time we have to define the basic individual-level processes that govern the transition of individuals from one compartment to another.

The simplest definition of epidemic dynamics considers the total population in the system as fixed, consisting of N individuals, and ignores any other demographic process (migrations, births, etc.). One of the simplest two-state compartmentalizations is the SIS model with only two possible transitions: The first one, denoted $S \to I$, occurs when a susceptible individual interacts with an infectious individual and becomes infected. The second transition, denoted $I \to S$, occurs when the infectious individual recovers from the disease and returns to the pool of susceptible individuals. The SIS model assumes that the disease does not confer immunity and individuals can be infected over and over again, undergoing a cycle $S \to I \to S$, which, under some conditions, can be sustained forever. Another basic model is the classic three-state SIR model. In the SIR model, the transition $I \to S$ of the SIS process is replaced by $I \to R$, which occurs when an infectious individual recovers from the disease and is assumed to have acquired a permanent immunity, or is removed (e.g., has died). Clearly, the SIR process does not allow for a stationary state.

The SIR and SIS models exemplify a basic classification of epidemic models given in terms of their long time behavior; see Fig. 1. In the long time regime, the SIS model can exhibit a stationary state, the *endemic* state, characterized by a constant (on average) fraction of infected individuals. In the SIR model, instead, the number of infected individuals always tends to zero.

In the SIS and SIR models, the infection and recovery processes completely determine the epidemic evolution. The $I \to R$ and $I \to S$ transitions occur spontaneously after a certain time the individuals spend fighting the disease or taking medical treatments; the transition does not depend on any interactions with other individuals in the population. The $S \to I$ transition instead occurs only because of the contact or interaction of the susceptible individual with an infectious one. In this case the interaction pattern among individuals is a specific feature of the transition and has to be taken into account.

The distribution of the "infectious period" and the transition probability can generally be estimated from clinical data. However, in a simplistic modeling scheme, the probability of transition is often assumed constant. In this way, a discrete-time formulation defines the recovery probability μ that

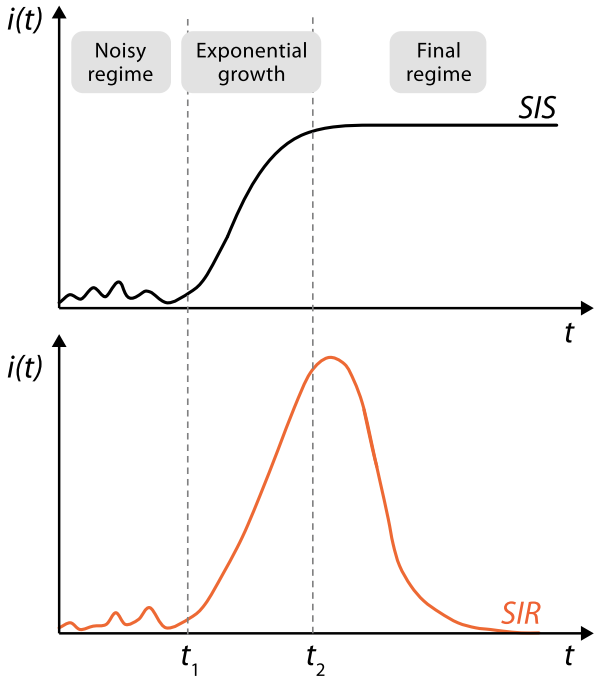

FIG. 1 (color online). Typical profile of the density $i(t)$ of infected individuals vs time in a given epidemic outbreak. In the first regime $t < t_1$, the outbreak is subject to strong statistical fluctuations. In the second regime $t_1 < t < t_2$, there is an exponential growth characterized by the details of the epidemic process. In the final regime ($t > t_2$), the density of infected individuals either converges to zero, for SIR-like models, or to a constant, possibly zero, for SIS-like models.

an individual will recover at any time step. The time an individual will spend on average in the infectious compartment, the mean infectious period, is then equal to μ^{-1} time steps. In a continuous-time formulation and assuming a Poisson process (Cox, 1967), μ is a rate (probability per unit time) and the probability that an individual remains infected for a time τ follows an exponential distribution $P_{\text{inf}}(\tau) = \mu e^{-\mu\tau}$, with an average infection time $\langle\tau\rangle = \mu^{-1}$. The Poisson assumption for the processes of infection and recovery leads naturally to a Markovian description of epidemic models (Ross, 1996).

The probability of the $S \to I$ transitions is more complicated and it is dependent on several factors and on the modeling approximations considered. In the absence of detailed data on human interactions, the most basic approach considers a homogenous mixing approximation (Anderson and May, 1992) which assumes that individuals interact randomly with each other. In this assumption, the larger the number of infectious individuals among an individual's contacts, the higher the probability of transmission of the infection. This readily translates to the definition of the force of infection α, that expresses the probability, also called the risk, at which one susceptible individual may contract the infection in a single time step. In the continuous-time limit we can define α as a rate and assume that

$$\alpha = \bar{\beta}\frac{N^I}{N}, \tag{1}$$

[1]In the rest of the review we will use the words removed or recovered in this context, interchangeably.

where $\bar{\beta}$ depends on the specific disease as well as the contact pattern of the population, and N^I is the number of infected individuals. Thus, α is proportional to the fraction $\rho^I = N^I/N$ of infected individuals in the population. In some cases $\bar{\beta}$ is explicitly split in two terms as βk, where β is now the rate of infection per effective contact and k is the number of contacts with other individuals. This form of the force of infection corresponds to the mass-action law (Hethcote, 2000), a widely used tool in the basic mean-field description of many dynamical processes in chemistry and physics. The force of infection depends only on the density of infectious individuals and decreases for larger populations, all other factors being equal. It is possible however to consider forces of infection of the type $\alpha = \beta N^I$, where the per capita infection probability is proportional to the actual number of infected individuals N^I, and assumes that the number of contacts scales proportionally to the size of the population. Indeed, also intermediate expressions for the force of infection depending on the size of the population as N^{-a} have been discussed in the literature (Anderson and May, 1992).

Generalizing the previous approach, an epidemic can be rephrased as a stochastic reaction-diffusion process (van Kampen, 1981). Individuals belonging to the different compartments can be represented as different kinds of "particles" or "species" that evolve according to a given set of mutual interaction rules, representing the different possible transitions among compartments, and that can be specified by means of appropriate stoichiometric equations. In the continous-time limit each reaction (transition) is defined by an appropriate *reaction rate*. We can therefore adopt the reaction-diffusion formalism to describe the basic epidemic models; see Fig. 2. The SIS model is thus governed by the reactions

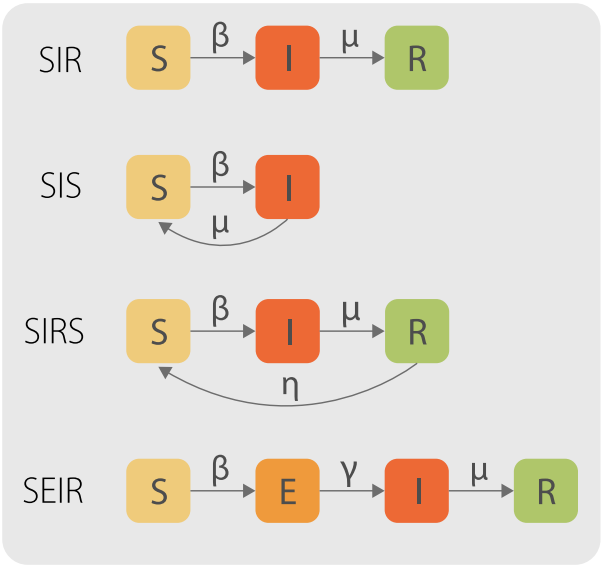

FIG. 2 (color online). Diagrammatic representation of different epidemic models in terms of reaction-diffusion processes. Boxes stand for different compartments, while the arrows represent transitions between compartments, happening stochastically according to their respective rates.

$$\begin{aligned} S + I &\xrightarrow{\beta} 2I, \\ I &\xrightarrow{\mu} S, \end{aligned} \tag{2}$$

where β and μ are transition rates for infection and recovery, respectively. In this model infection can be sustained forever for sufficiently large β or small μ. The SIR model (Kermack and McKendrick, 1927) is instead characterized by the three compartments S, I, and R, coupled by the reactions

$$\begin{aligned} S + I &\xrightarrow{\beta} 2I, \\ I &\xrightarrow{\mu} R. \end{aligned} \tag{3}$$

For any values of β and μ, the SIR process will always asymptotically die after affecting a given fraction of the population.

Many more epidemic models can be defined analogously to the SIS and SIR models. A useful variant is the SI model, which considers only the first transition in Eqs. (2) and (3), i.e., individuals become infected and never leave this state. While the SI model is a somewhat strong simplification (valid only in cases where the time scale of recovery is much larger than the time scale of infection), it approximates the initial time evolution of both SIS and SIR dynamics. More realistic models are defined in order to better accommodate the biological properties of real diseases. For instance, the susceptible-infected-recovered-susceptible (SIRS) model is an epidemic model incorporating a temporary immunity. It can be defined from the SIR model by adding a microscopic transition event

$$R \xrightarrow{\eta} S, \tag{4}$$

where η is the rate at which the immunity of a recovered individual is lost, rendering him or her susceptible again. The SEIR model is a variation of the SIR model including the effects of exposed (E) individuals, which have been infected by the disease but cannot yet transmit it. The SEIR model is one of the paradigmatic models for the spreading of influenza-like illnesses and in the compact reaction-diffusion notation reads as

$$\begin{aligned} S + I &\xrightarrow{\beta} E + I, \\ E &\xrightarrow{\gamma} I, \\ I &\xrightarrow{\mu} R. \end{aligned} \tag{5}$$

All the above models can be generalized to include demographic effects (birth and death processes in the population), the age structure of the population, other relevant compartments (such as asymptomatic infected individuals), etc. A more complete and detailed review of epidemic models and their behavior can be found in Anderson and May (1992), Keeling and Rohani (2007), and Brauer and Castillo-Chavez (2010).

B. Basic results from classical epidemiology

Although epidemic spreading is best described as a stochastic reaction-diffusion process, the classic understanding

of epidemic dynamics is based on taking the continuous-time limit of difference equations for the evolution of the average number of individuals in each compartment. This deterministic approach relies on the homogeneous mixing approximation, which assumes that the individuals in the population are well mixed and interact with each other completely at random, in such a way that each member in a compartment is treated similarly and indistinguishably from the others in that same compartment. This approximation, which is essentially equivalent to the mean-field approximation commonly used in statistical physics, for both equilibrium (Stanley, 1971) and nonequilibrium (Marro and Dickman, 1999) systems, can be shown to be correct in regular lattices with high dimension, but it is not exact in low dimensions (ben-Avraham and Havlin, 2005). Under this approximation, full information about the state of the epidemics is encoded in the total number N^α of individuals in the compartment α or, analogously, in the respective densities $\rho^\alpha = N^\alpha/N$, where N is the population size. The time evolution of the epidemics is described by deterministic differential equations, which are constructed applying the law of mass action, stating that the average change in the population density of each compartment due to interactions is given by the product of the force of infection times the average population density (Hethcote, 2000).

The deterministic equations for the SIR and SIS processes are obtained by applying the law of mass action and read as

$$\frac{d\rho^I}{dt} = \beta\rho^I\rho^S - \mu\rho^I, \tag{6}$$

$$\frac{d\rho^S}{dt} = -\beta\rho^I\rho^S + \chi\rho^I, \tag{7}$$

where $\chi = \mu$ for the SIS process and $\chi = 0$ for the SIR model, and the force of infection is $\alpha = \beta\rho^I$. These equations are complemented with the normalization conditions $\rho^R = 1 - \rho^S - \rho^I$ and $\rho^S = 1 - \rho^I$ for the SIR and SIS models, respectively. If we consider the limit $\rho^I \simeq 0$, generally valid at the early stage of the epidemic, we can linearize the above equations obtaining for both the SIS and SIR models the simple equation

$$\frac{d\rho^I}{dt} \simeq (\beta - \mu)\rho^I \tag{8}$$

whose solution

$$\rho^I(t) \simeq \rho^I(0)e^{(\beta-\mu)t} \tag{9}$$

represents the early time evolution. Equation (9) illustrates one of the key concepts in the classical theoretical analysis of epidemic models. The number of infectious individuals grows exponentially if

$$\beta - \mu > 0 \Rightarrow R_0 = \frac{\beta}{\mu} > 1, \tag{10}$$

where we defined the basic reproduction number R_0 as the average number of secondary infections caused by a primary case introduced in a fully susceptible population (Anderson and May, 1992). This result allows one to define the concept of epidemic threshold: only if $R_0 > 1$ (i.e., if a single infected individual generates on average more than one secondary infection) an infective agent can cause an outbreak of a finite relative size (in SIR-like models) or lead to a steady state with a finite average density of infected individuals, corresponding to an endemic state (in SIS-like models). If $R_0 < 1$ (i.e., if a single infected individual generates less than one secondary infection), the relative size of the epidemics is negligibly small, vanishing in the thermodynamic limit of an infinite population[2] (in SIR-like models) or leading to a unique steady state with all individuals healthy (in SIS-like models). This concept is very general and the analysis of different epidemic models (Anderson and May, 1992) reveals in general the presence of a *threshold behavior*, with a reproduction number that can be expressed as a function of the rates of the different transitions describing the epidemic model.

A few remarks are in order here. First, although we have stated that epidemic processes can be considered as reaction-diffusion systems, the classic approach completely neglects the diffusion of individuals. Spatial effects can be introduced by adding diffusive continuous terms or by considering patch models. Furthermore, epidemic spreading is governed by an inherently probabilistic process. Therefore, a correct analysis of epidemic models should consider explicitly its stochastic nature (Andersson and Britton, 2000). Accounting for this stochasticity is particularly important when dealing with small populations, in which the number of individuals in each compartment is reduced. For instance, while the epidemic threshold condition $R_0 > 1$ is a necessary and sufficient condition for the occurrence of an epidemic outbreak in deterministic systems, in stochastic systems this is just a necessary condition. Indeed even for $R_0 > 1$ stochastic fluctuations can lead to the epidemic extinction when the number of infectious individuals is small. Analogously, all the general results derived from deterministic mean-field equations can be considered representative of real systems only when the population size is very large (ideally in the thermodynamic limit) and the fluctuations in the number of individuals can be considered small. Indeed, most of the classical results of mathematical epidemiology have been obtained under these assumptions (Anderson and May, 1992).

Another point worth stressing is the Poisson assumption. Although we mostly focus on Poissonian epidemic processes (see Secs. VII.A and VIII for some remarks on the non-Poissonian case), a different phenomenology, both more complex and interesting, can be obtained from nonexponentially distributed infection or recovery processes.

Finally, the classic deterministic approach assumes random and homogeneous mixing, where each member in a compartment is treated similarly and indistinguishably from the others in that same compartment. In reality, however, each individual has his or her own social contact network over which diseases propagate, usually differing from that of other members in a group or compartment. Diekmann, Heesterbeek, and Britton (2012) illustrated the weakness of R_0 by discussing a line and square lattice topology and they concluded that network and

[2]In the present context, since we do not consider spatial effects, the thermodynamic limit is simply defined as the limit of an infinitely large number of individuals.

percolation theory needs to be consulted to compute the epidemic threshold, leading to a new definition of the basic reproduction number depending on the topology of the network. Thus, for example, in the case of a homogeneous contact network in which every individual is in contact with the same number of individuals $\langle k \rangle$, the basic reproduction number takes the form

$$R_0 = \langle k \rangle \frac{\beta}{\mu}. \tag{11}$$

The impact of heterogeneous connectivity patterns, reflected by an underlying network topology, on the epidemic behavior is the focus of the present review.

C. Connections with statistical physics models

The interest that models for epidemic spreading have attracted within the statistical physics community stems from the close connection between these models and more standard nonequilibrium problems in statistical physics (Marro and Dickman, 1999; Henkel, Hinrichsen, and Lübeck, 2008). In particular, the epidemic threshold concept is analogous to the concept of phase transition in nonequilibrium systems. A phase transition is defined as an abrupt change in the state (phase) of a system, characterized by qualitatively different properties, and that is experienced varying a given control parameter λ. The transition is characterized by an order parameter ρ (Yeomans, 1992), which takes (in a system of infinite size) a nonzero value in one phase, and a zero value in another (see Fig. 3). The phase transition takes place at a particular value of the control parameter, the so-called transition point λ_c, in such a way that for $\lambda > \lambda_c$ we have $\rho > 0$, while for $\lambda \leq \lambda_c$, $\rho = 0$. Apart from the determination of the transition point, the interest in physics lies in the behavior of the order parameter around λ_c, which in continuous or critical phase transitions[3] takes a power-law form $\rho(\lambda) \sim (\lambda - \lambda_c)^{\beta_{\text{crit}}}$, defining the critical exponent β_{crit} (Yeomans, 1992).

The SIS dynamics thus belongs to the wide class of nonequilibrium statistical models possessing absorbing states, i.e., states in which the dynamics becomes trapped with no possibility to escape. The paradigmatic example of a system with an absorbing state is the contact process (CP) (Harris, 1974), where all nodes of a lattice or network can be either occupied or empty. Occupied nodes annihilate at rate 1; on the other hand, they can reproduce at rate λ, generating one offspring that can occupy an empty nearest neighbor. The contact process experiences an absorbing state phase transtion (Marro and Dickman, 1999; Henkel, Hinrichsen, and Lübeck, 2008) at a critical point λ_c between an active phase, in which activity lasts forever in the thermodynamic limit, implying a finite average density of occupied nodes, and an absorbing phase, in which activity eventually vanishes, corresponding to an empty system. In the case of the SIS model, the active phase is given by the infected state, and the absorbing phase by the state where no individual is infected; see Fig. 3. The order parameter is therefore the prevalence or density of infected individuals, and the control parameter is given by the spreading rate or effective infection rate, which equals $\lambda = \beta/\mu$. The epidemic threshold (critical point) λ_c separates thus the infected from the healthy phase. While this distinction is strictly true in the thermodynamic limit, for finite systems the dynamics for any value of λ sooner or later visits the absorbing state and remains trapped there. The absorption event can occur even in the active phase well above the critical point, because of random fluctuations, illustrating that the determination of the critical point is a nontrivial task, for both theoretical approaches and numerical simulations (Marro and Dickman, 1999; Henkel, Hinrichsen, and Lübeck, 2008). It is interesting to note that the dynamics of the SIS process is essentially identical to that of the contact process in lattices; indeed, the difference between the SIS and the contact process lies exclusively in the number of offsprings that an active individual can generate. While in the contact process one particle generates always on average one offspring per unit time, an infected individual in the SIS model can infect all his or her nearest neighbors in the same time interval. This difference is trivial when the number of nearest neighbors is fixed, but it can lead to a dramatic difference when the number of nearest neighbors has large fluctuations (see Sec. V).

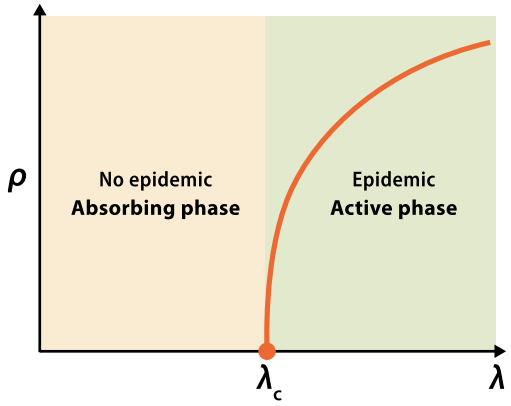

FIG. 3 (color online). Phase diagram of a typical nonequilibrium absorbing state phase transition (SIS-like). Below the critical point λ_c, the order parameter is zero (healthy phase in an epidemics interpretation). Above the critical point, the order parameter attains a nonzero average value in the long time regime (endemic or infected epidemic phase).

The SIR model also exhibits a transition between a phase where the disease outbreak reaches a finite fraction of the population and a phase where only a limited number of individuals are affected. This is strongly reminiscent of the transition occurring in percolation (Grassberger, 1983; Stauffer and Aharony, 1994). In the simplest possible setting of (bond) percolation in a lattice, the connections between nearest neighbors of a lattice or network are erased with probability $1 - p$ and kept with complementary probability p. A critical value p_c separates a supercritical percolating phase, where a macroscopic connected cluster spans the whole lattice, from a subcritical phase where only connected clusters of finite size exist. The order parameter describing the transition is the probability $P_G(p)$ that a randomly chosen

[3]In first-order transitions the order parameter takes a discontinuous jump at the transition point (Stanley, 1971).

site belongs to the spanning cluster. In the case of networks, the percolating phase corresponds to the presence of the largest connected component with a size proportional to the network size (the giant component, see Sec. III.A), while in the subcritical phase it has a relative size that vanishes in the thermodynamic limit. In the case of networks, the order parameter is proportional to the relative size of the giant component. The mapping between SIR and bond percolation is made through the assimilation of the size of connected components with the size of epidemic outbreaks, with a control parameter that depends on the spreading rate $\lambda = \beta/\mu$. This connection will be further developed and exploited in Sec. V.B.

Finally, it is worth mentioning first-passage percolation (Hammersley and Welsh, 1965; Kesten, 2003) as another classical problem related to epidemics. In this model, a nonnegative value τ_{ij} is defined on each edge of a graph and interpreted as the time needed to cross the edge. Given a topology and the distribution of the times τ_{ij}, first-passage percolation investigates which points can be reached in a certain time starting from a fixed origin. The SI model for epidemics can be seen as the limit of first-passage percolation with all passage times distributed exponentially.

III. NETWORK MEASURES AND MODELS

Although very common, the homogeneous assumption used in the previous section to derive the constitutive deterministic equations of basic epidemic processes may be inadequate in several real-world situations where individuals have large heterogeneity in the contact rate, a specific frozen pattern of interaction, or are in contact with only a small part of the population. These features may have different relevance depending on the disease or contagion process considered. However, a wide range of social and biological contagion processes requires capturing the individuals' contact pattern structure in the mathematical modeling approaches. This is even more relevant, because most real-world systems show very complex connectivity patterns dominated by large-scale heterogeneities described by heavy-tailed statistical distributions.

Network theory (Newman, 2010) provides a general framework to discuss interactions among individuals in detail. In this section, we provide a short summary of the main definitions and properties of networks, relevant for epidemic spreading, and a basic introduction to the language of graph theory that is necessary for a formal analysis of network properties. Network science is burgeoning at the moment, and for more extensive accounts of this field see Dorogovtsev and Mendes (2003), Caldarelli (2007), Cohen and Havlin (2010), Dorogovtsev (2010), Newman (2010), and Barabási (2015).

A. General definitions

Networks are mathematically described as graphs. A graph is a collection of points, called vertices (nodes in the physics literature or actors in the social sciences). These points are joined by a set of connections, called edges, links, and ties, in mathematics, physics, and social sciences, respectively. Each edge denotes the presence of a relation or interaction between the vertices it joins. Edges can represent a bidirectional interaction between vertices, or indicate a precise directionality in the interaction. In the first case we talk about undirected networks, and in the second case, about directed networks or digraphs. From an epidemiological point of view, the directedness of a network is indeed relevant since it imposes restrictions on the possible paths of propagation of the contagion. A compact way to specify all connections present in a graph of size N (i.e., with N vertices) is the $N \times N$ adjacency matrix A, with elements $a_{ij} = 1$ if an edge is connecting nodes i and j and zero otherwise. A is symmetric in undirected graphs and asymmetric in directed graphs.

A path $\mathcal{P}_{i_0,i_n}$ connecting vertices i_0 and i_n is a sequence of connected edges $\{(i_j, i_{j+1})\}$, $j = 0, \ldots, n-1$; the number of edges traversed n is the hop count, also called the length, of the path. A graph is connected if there exists a path connecting any two vertices in the graph. A loop is a closed path with $i_0 \equiv i_n$. A component $\mathcal{C}$ of a graph is defined as a connected subgraph. The giant component is the component or subgraph, whose size scales as the number of vertices in the graph. From an epidemiological perspective, a disease in the giant component may in principle infect a macroscopic fraction of the graph, while if the disease starts outside of the giant component, the total number of infected vertices will be necessarily limited, representing a fraction that decreases with the network size.

In the case of directed graphs, the structure of the components is more complex as the presence of a path from node i to node j does not necessarily guarantee the presence of a corresponding path from j to i. In general (see Fig. 4) the component structure of a directed network can be decomposed into a giant weakly connected component (GWCC), corresponding to the giant component of the same graph in which the edges are considered as undirected, plus a set of smaller disconnected components. The GWCC is itself composed of several parts because of the directed nature of its edges: (1) the giant strongly connected component (GSCC), in which there is a directed path joining any pair of nodes; (2) the giant in component (GIN), formed by the nodes from which it is possible to reach the GSCC by means of a directed

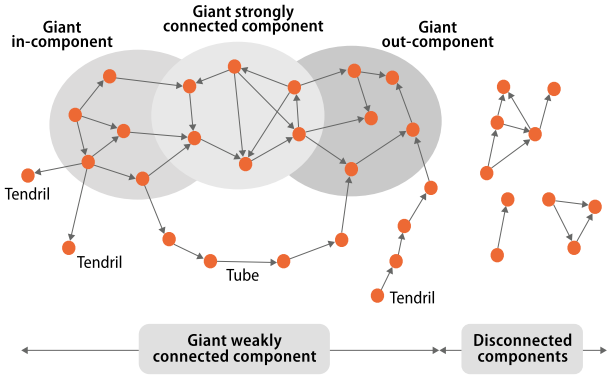

FIG. 4 (color online). Component structure of a directed graph. Adapted from Dorogovtsev, Mendes, and Samukhin, 2001.

path; (4) the tendrils that connect nodes that cannot reach the GSCC or be reached from it; and (5) the tubes that connect the GIN and GOUT, but do not belong to the GSCC.

B. Network metrics

A large number of metrics have been defined to characterize different aspects of the topology of complex networks.

1. Shortest path length and network diameter

In order to characterize the distance among nodes we introduce the shortest path length, sometimes also referred to as the chemical distance or geodesical distance. The shortest path distance ℓ_{ij} between two nodes i and j is defined as the length of the shortest path (not necessarily unique) joining i and j. The diameter of a network is the maximum value of all the pairwise shortest path lengths, and the average shortest path length $\langle \ell \rangle$ is the average of the value of ℓ_{ij} over all pairs of vertices in the network.

2. Degree and degree distribution

The degree k_i of vertex i in an undirected network is the number of edges emanating from i, i.e., $k_i = \sum_j a_{ij}$. In the case of directed networks, we distinguish between in-degree k_i^{in} and out-degree k_i^{out} as the number of edges that end in i or start from i, respectively. In undirected networks we define the degree distribution $P(k)$ as the probability that a randomly chosen vertex has degree k, or, in finite networks, as the fraction of vertices in the graph with degree exactly equal to k. In the case of directed networks, there are instead two different distributions, the out-degree $P_{\text{out}}(k^{\text{out}})$ and in-degree $P_{\text{in}}(k^{\text{in}})$ distributions. The in-degree and out-degree of a given vertex might be not independent. Correlations are encoded in the joint probability distribution $P(k^{\text{in}}, k^{\text{out}})$ that a randomly chosen vertex has in-degree k^{in} and out-degree k^{out}. It is useful to consider the moments of the degree distribution $\langle k^n \rangle = \sum_k k^n P(k)$. The first moment, the average degree $\langle k \rangle = 2L/N$, twice the ratio between the number L of edges (or links) and the number N of nodes, provides information about the density of the network. A network is called sparse if its number of edges L grows at most linearly with the network size N; otherwise, it is called dense. In directed networks, since every edge contributes to the in-degree of one node and to the out-degree of another node we have that $\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle$.

3. Degree correlations

Two-vertex degree correlations can be conveniently measured by means of the conditional probability $P(k'|k)$ that an edge departing from a vertex of degree k is connected to a vertex of degree k' (Pastor-Satorras, Vázquez, and Vespignani, 2001). A network is called uncorrelated if this conditional probability is independent of the originating vertex k. In this case, $P(k'|k)$ can be simply estimated as the ratio between the number of edges pointing to vertices of degree k', $k'P(k')N/2$, and the total number of edges, $\langle k \rangle N/2$, to yield $P^{\text{un}}(k'|k) = k'P(k')/\langle k \rangle$. The empirical evaluation of $P(k'|k)$ turns out to be quite noisy in real networks, due to finite-size effects. A related, simpler, measure of correlations is the average degree of the nearest neighbors of vertices of degree k, $\bar{k}_{nn}(k)$ which is formally defined as (Pastor-Satorras, Vázquez, and Vespignani, 2001)

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k). \tag{12}$$

For uncorrelated networks, $\bar{k}_{nn}^{\text{un}}(k) = \langle k^2 \rangle / \langle k \rangle$ does not depend on k. Therefore, a varying $\bar{k}_{nn}(k)$ is the signature of degree correlations. The analysis of empirical networks has suggested a broad classification of networks in two main classes, according to the nature of their degree correlations (Newman, 2002a): Assortative networks exhibit an increasing $\bar{k}_{nn}(k)$, indicative that high degree nodes tend to connect to high degree nodes, while low degree nodes are preferentially attached to low degree nodes. Disassortative networks, on the other hand, show a decreasing $\bar{k}_{nn}(k)$ function, suggesting that high degree nodes connect to low degree nodes, and vice versa. Assortativity by degree can be characterized by the Pearson correlation coefficient r (Newman, 2002a): Uncorrelated networks have $r = 0$, while assortative (disassortative) networks present $r > 0$ ($r < 0$), respectively.

4. Clustering coefficient and clustering spectrum

The concept of clustering refers to network transitivity, i.e., the relative propensity of two nodes to be connected, provided that they share a common neighbor. The clustering coefficient C is defined as the ratio between the number of loops of length three in the network (i.e., triangles), and the number of connected triples (three nodes connected by two edges). A local measure c_i of clustering (Watts and Strogatz, 1998) can also be defined as the ratio between the actual number of edges among the neighbors of a vertex i, e_i, and its maximum possible value, thus directly measuring the probability that two neighbors of vertex i are also neighbors of each other. The mean clustering of the network $\langle c \rangle$ is defined as the average of c_i over all vertices in the network. The clustering spectrum $\bar{c}(k)$ is defined as the average clustering coefficient of the vertices of degree k (Vázquez, Pastor-Satorras, and Vespignani, 2002; Ravasz and Barabási, 2003), satisfying $\langle c \rangle = \sum_k P(k)\bar{c}(k)$.

5. Centrality and structure in networks

The concept of centrality encodes the relative importance of a node inside a network, a relevant issue in the context of social network analysis (Wasserman and Faust, 1994). Many different definitions of centrality have been proposed, based on different indicators of the structural importance of nodes. The simplest of them is the degree, referred to as degree centrality. The higher its degree, the more the node can be considered influential and/or central in the network. Alternative definitions are based on the shortest paths between vertices. Thus, the closeness centrality $\mathcal{C}_i$ is defined as the inverse of the average of the shortest path lengths from vertex i to all other vertices in the network. With this measure, we consider a vertex central if it is situated on average at a short distance to all other vertices in the network. A different perspective on centrality is provided by the betweenness centrality b_i of vertex i, defined as the number of shortest

paths between any two vertices in the network that pass through vertex i. More precisely, if $L_{h,j}$ is the total number of shortest paths from h to j, and $L_{h,i,j}$ is the number of these shortest paths that pass through vertex i, then $b_i = \sum_{h \neq j} L_{h,i,j}/L_{h,j}$. Betweenness measures thus centrality from the perspective of the control of information flowing between different nodes, assuming this information flows following the shortest path route (Freeman, 1977).

Another way to characterize the centrality of nodes resides in the concept of K-coreness. The K-core of a network is a maximal connected subgraph, such that all vertices in the subgraph have degree $k \geq K$ (Seidman, 1983). The K-core decomposition is an iterative procedure that classifies the vertices of the network in nested levels of increasing connectivity (increasing K-core). The algorithm runs as follows: One starts with the complete network and removes iteratively all vertices with degree $k = 1$, until only vertices with degree $k \geq 2$ are present. The set of removed nodes represents the $K = 1$-shell, while the remaining nodes constitute the $K = 2$-core. In the next iteration of the process, all vertices with degree $k = 2$ are removed (the $K = 2$-shell), and we are left with the $K = 3$-core. This iterative process is stopped when we arrive at the maximum K_S-core, where one more application of the algorithm leaves no vertices. At each node is assigned a centrality measure equal to its K-core index, the deeper the more central.

It is worth remarking that real networks can display higher levels of architecture that are difficult to capture with a single number. Many networks possess a community structure, in which different sets of nodes, called communities or modules, have a relatively high density of internal connections, while they are more loosely connected among them. The problem of computing the community structure of a given network has been an active topic in network science and a large number of different approaches have been considered [see Fortunato (2010) for a review].

C. Generalizations of simple graphs

The simple concept of graph considered above can be refined at different levels, adding more and more complexity and detail in order to better represent the real system under consideration. A first extension is that of bipartite graphs, in which we have two different kinds of nodes, and edges join only two nodes of a different kind. A classical example is the network of heterosexual sexual relationships (Liljeros *et al.*, 2001).

Another important generalization consists of the definition of weighted networks, in which a real number ω_{ij} (the weight) is associated with the edge between vertices i and j. Weighted networks constitute the natural choice to represent many systems, including transportation networks (e.g., the airport network), in which the weight of an edge measures the fraction of people or goods transported by the edge in a given interval of time, or social networks, for which weights measure the relative intensity or frequency of contacts between pairs of vertices. The addition of weights allows one to define a complete new set of topological metrics (Braunstein *et al.*, 2003; Barrat, Barthélemy, Pastor-Satorras, and Vespignani, 2004; Onnela *et al.*, 2005; Serrano, Boguñá, and Pastor-Satorras, 2006; Ahnert *et al.*, 2007). Among those, the strength of a node s_i, defined as the sum of the weights of all edges incident to it, i.e., $s_i = \sum_j \omega_{ij}$, generalizes to weighted networks the concept of degree.

D. Network classes and basic network models

The recent abundance of data and measurements of real-world networks has highlighted the existence of different classes of networks, characterized by a large variability in basic metrics and statistical properties. This classification in its turn has fueled an intense theoretical research effort devoted to the study of different network generation models. The usefulness of these models in the present context is that they serve as generators of synthetic networks, with controlled topological properties, in which the behavior of dynamical processes such as epidemics can be studied in detail. In the following we survey some of the main network classes and models that are used for exploring the properties of epidemic processes.

1. Random homogenous networks

The first theoretical model of random networks is the classical random graph model (Solomonoff and Rapoport, 1951; Erdős and Rényi, 1959; Gilbert, 1959). In its simplest formulation, the graph $G_p(N)$ is constructed from a set of N nodes in which each one of the $N(N-1)/2$ possible links is present with probability p. The degree distribution is given by a binomial form, which in the limit of constant average degree (i.e., $p = \langle k \rangle / N$) and large N can be approximated by a Poisson distribution $P(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$. The clustering coefficient is simply given by $\langle c \rangle = p$, and the average shortest path length is $\langle \ell \rangle \simeq \log N / \log \langle k \rangle$ (Dorogovtsev, 2010). This model is therefore adequate in the case of networks governed only by stochasticity, although $G_p(N)$ tends to a regular graph for large N and constant p. The degree distribution is peaked around the average value, thus denoting a statistical homogeneity of the nodes. Interestingly, the model features for $\langle k \rangle > 1$ the small diameter observed in most real-world networks. However, any other structural properties, including the generally high clustering coefficient observed in real-world networks, cannot be reproduced by this model.

2. Small-world networks

The small-world model of Watts and Strogatz (1998) represents a first attempt to obtain a network with small diameter $\langle \ell \rangle$ and large clustering coefficient. This model considers an ordered lattice, such as a ring of N vertices, each one of which is symmetrically connected to its $2m$ nearest neighbors. This initial configuration has a large clustering coefficient and large average shortest path length. Starting from it, a fraction p of edges in the network is rewired, by visiting all m clockwise edges of each vertex and reconnecting them, with probability p, to a randomly chosen node. In another version of the model (Monasson, 1999), a fraction p of edges is added between randomly chosen pairs of vertices. The overall effect of the rewiring processes is to add long-range shortcuts, that, even for a small value of $p \sim N^{-1}$,

greatly reduce the average shortest path length, while preserving a large clustering for not very large values of p. This model, although better suited for social networks with high clustering coefficient, has a degree distribution and centrality measures decaying exponentially fast away from the average value. The small-world model thus generates homogeneous networks where the average of each metric is a typical value shared, with little variations, by all nodes of the network.

3. Heavy-tailed networks

Empirical evidence from different research areas has shown that many real-world networks exhibit levels of heterogeneity not anticipated until a few years ago. The statistical distributions characterizing heterogeneous networks are generally skewed and varying over several orders of magnitude. Thus, real-world networks are structured in a hierarchy of nodes with a few nodes having very large connectivity (the hubs), while the vast majority of nodes have much smaller degrees. More precisely, in contrast with regular lattices and homogeneous graphs characterized by a typical degree k close to the average $\langle k \rangle$, heterogeneous networks exhibit heavy-tailed degree distributions often approximated by a power-law behavior of the form $P(k) \sim k^{-\gamma}$, which implies a non-negligible probability of finding vertices with very large degree. The degree exponent γ of many real-world networks takes a value between 2 and 3. In such cases networks are called scale free, since the second moment of the degree distribution diverges in the infinite network size limit ($N \to \infty$). It is understood that in real-world networks the finite size N and the presence of biological, cognitive, and physical constraints impose an upper limit to the second degree moment. However, the second moment of the distribution is in many cases overwhelmingly large, reflecting enormous connectivity fluctuations. The presence of large-scale fluctuations associated with heavy-tailed distributions is often true not only for the degree of nodes but it is also observed for the intensity carried by the connecting links, transport flows, and other basic quantities.

Several variations of the classical random graph model have been proposed in order to generate networks with a power-law degree distribution. One variation, the so-called configuration model (Bender and Canfield, 1978; Molloy and Reed, 1995), considers a random network with a fixed degree distribution, instead of the fixed average degree of classical random graphs. Its construction is as follows: To each of the vertices, we assign a degree k_i, given by a random number selected from the probability distribution $P(k)$, subject to the conditions $m \leq k_i \leq N$, where m is the desired minimum degree, and such that $\sum_i k_i$ is an even number. The actual graph is constructed by randomly connecting the nodes with $\sum_i k_i/2$ edges, preserving the degree originally assigned. In finite networks, an average maximum degree or degree cutoff k_m, known as the natural cutoff of the network (Boguñá, Pastor-Satorras, and Vespignani, 2004) is observed, which is a function of the network size of the form $k_m(N) \sim N^{1/(\gamma-1)}$ (Cohen *et al.*, 2000). The original configuration model leads for power-law distributions with $\gamma \leq 3$ to the formation of networks with multiple and self-connections. The additional prescription that multiple and self-connections are removed leads to the generation of disassortative correlations (Park and Newman, 2003; Maslov, Sneppen, and Zaliznyak, 2004). These correlations are avoided in the uncorrelated configuration model (Catanzaro, Boguñá, and Pastor-Satorras, 2005) by imposing a hard structural cutoff $k_m \sim N^{1/2}$.

A different modeling paradigm, namely, the class of growing network models, is based on the empirical observation that many real networks do not have a constant number of vertices and edges, but are instead growing entities, in which nodes and links are added over time. The first undirected model of this kind is the Barabási-Albert model (Barabási and Albert, 1999), based on the assumption that newly added edges will tend in general to be connected to nodes chosen via some preferential attachment rule. The simplest of these preferential rules is a degree-biased rule, in which the probability to add a connection to a vertex i is some function $F(k_i)$ of its degree. The Barabási-Albert model, assuming the simplest, linear, form for the preferential attachment function, is defined as follows: (i) The network starts with a small nucleus of m_0 connected vertices; every time step a new node is added, with m ($m \leq m_0$) edges which are connected to old vertices in the network. (ii) New edges are connected to the ith node in the network with probability equal to $F(k_i) = k_i / \sum_j k_j$. In the long time limit, the network thus generated has a degree distribution $P(k) \sim k^{-3}$ (Barabási and Albert, 1999; Dorogovtsev, Mendes, and Samukhin, 2000). The original growing network model has been subject to an impressive number of variations and extensions toward realistic growing dynamics and to accommodate for different exponents of the degree distribution and other properties such as high clustering and tunable degree-degree correlations (Newman, 2010).

E. Static versus dynamic networks

So far we assumed that the topology defining the network is static: the set of nodes and links do not change over time. However, many other real networks are far from static, their links being created, destroyed, and rewired at some intrinsic time scales. In some of these dynamical networks, such as the Internet (Pastor-Satorras and Vespignani, 2004), the time scale of the network evolution is quite slow. A static network provides a good approximation, when the properties of dynamical processes evolve at a much faster time scale than topological changes. The opposite limit defines the so-called annealed networks (Gil and Zanette, 2005; Stauffer and Sahimi, 2005; Weber and Porto, 2007; Boguñá, Castellano, and Pastor-Satorras, 2009), which describe the case when the evolution of the network is much faster than the dynamical processes. In this limit, the dynamical process unfolds on a network that is rapidly rewiring so that the dynamics effectively occurs on an average network in which each connection is possible according to a specific probability that depends on the degree distribution $P(k)$ and the two-node degree correlations $P(k'|k)$. An annealed network is thus described by a mean-field version of the adjacency matrix that is presented in Sec. IV.

The above two limits are relevant in the definition of the approximations and the limits of applicability of the most

commonly used theoretical approaches to epidemic spreading in networks. There are, however, several other instances of networks, such as in social systems, where the connectivity pattern varies over time scales comparable to those of the dynamical processes on top of it and it is crucial to explicitly take into account the concurrent dynamics of the spreading process and the connectivity pattern. The effect on epidemic spreading of the dynamical nature of such temporal (Holme and Saramäki, 2012) networks is discussed in Sec. VIII.

Finally, coevolution of the network and the dynamical process occurs when the topological structure of a network reacts dynamically to the evolution of a dynamical process taking place on top of it. Indeed, individual social activity can be altered by the presence of an epidemic outbreak (e.g., avoiding contacts that amount to link deletion), thus affecting the topology of the underlying social network, which in turn feeds back nontrivially on the spreading dynamics. The coupling of topology with disease evolution in such coevolving networks is discussed in Sec. VII.B.7.

IV. THEORETICAL APPROACHES FOR EPIDEMIC MODELING ON NETWORKS

A continuous-time epidemic process with constant transition rates between compartments on any graph can be described by Markov chain theory. We consider a network defined by its adjacency matrix A and a general epidemic process with q compartments. The state of node i at time t is specified by a random variable $X_i(t) \in \{0, 1, \ldots, q-1\}$, where $X_i(t) = \alpha$ means that node i belongs to compartment α at time t. We assume that all transitions between compartments are given by independent Poisson processes with given rates. Under these conditions, the evolution of the epidemic process can be described in terms of a Markov chain (van Kampen, 1981; Van Mieghem, 2014b). In a network with N nodes, the total number of states equals q^N, all possible combinations in which all N nodes can take a value from 0 to $q-1$. The elements of the $q^N \times q^N$ infinitesimal generator Q of the continuous-time Markov chain are explicitly computed for $q = 2$ in Van Mieghem, Omic, and Kooij (2009), Simon, Taylor, and Kiss (2011), and Van Mieghem and Cator (2012), while the general case is treated in Darabi Sahneh, Scoglio, and Van Mieghem (2013). Once the infinitesimal generator Q and the initial infection probabilities are known, the state probabilities $\Pr[X_1(t) = x_1, \ldots, X_N(t) = x_N]$ at time t, for each $x_j = 0, 1, \ldots, q-1$, can be computed using ordinary matrix operations, from which all desired information can be deduced in principle.

Although the Markov approach is exact, its use has been limited to a few exact results in the case of the SIS model. Indeed, using an exact Markov approach is impervious for a number of reasons. First, the linear set of $q^N \times q^N$ equations to be solved limits the analysis to very small graphs. Second, the structure of the infinitesimal generator Q is rather complex, which prevents one from gaining general insights, although it is possible (Van Mieghem and Cator, 2012) to deduce a recursion relation between the Q matrix in a graph with N and $N+1$ nodes. Third, in most cases, we are interested in the steady-state (or stationary) behavior or in the final size of the epidemic. The peculiar property of the exact continuous-time Markov process is the appearance of an absorbing state, which is equal to the overall-healthy state ($x_j = 0$ for each node j) in which the activity (virus, information spreading, etc.) has disappeared from the network. Mathematically, an absorbing state means that the Q matrix has a row of zero elements, the Markov chain is reducible, and the steady state is equal to this overall-healthy state for finite N. These complications mean that only a time-dependent analysis, focusing on metastable states, may answer questions of practical interest.

More, in general, few exact results have been derived for epidemic spreading in networks. For this reason, the derivation of explicit results on the behavior of epidemic spreading processes in networks mostly relies on mean-field theoretical approaches of a different kind. In the following we review these approaches and discuss the different approximations and assumptions on which they are based. The detailed applications of these approaches to the paradigmatic cases of the SIS and SIR models will be presented in Sec. V.

A. Individual-based mean-field approach

Individual-based mean-field (IBMF) theory represents a drastic simplification of the exact description presented previously. The basic idea (Wang *et al.*, 2003; Chakrabarti *et al.*, 2008; Van Mieghem, Omic, and Kooij, 2009; Gómez *et al.*, 2010) is to write down evolution equations for the probability ρ_i^α that the node i belongs to the compartment α, for any node i, assuming that the dynamic state of every node is statistically independent of the state of its nearest neighbors. The mean-field equations can be obtained, under this assumption, by applying an extended version of the law of mass action, i.e., assuming that the probability that node i is in state α and its neighbor node j in state α' is $\rho_i^\alpha \rho_j^{\alpha'}$. More systematically, they can be obtained directly from the governing equations derived from the q^N-state Markov chain, assuming that the expected values of variable pairs factorize $E[X_i X_j] = E[X_i]E[X_j]$. This method is akin to the classic assumption of the mean-field theory, while keeping the full topological structure of the network encoded in all the entries of the adjacency matrix a_{ij}, that it is considered to be static or quenched, using the language of mean-field theory in statistical mechanics.

The solutions of IBMF theories depend in general on the spectral properties of the adjacency matrix, and, in particular, on the value of its largest eigenvalue Λ_1. Their predictions are generally in agreement with numerical simulation results obtained for static networks. As is well known from the theory of critical phenomena, the agreement tends to decrease, when the densities $\rho_i^\alpha \to 0$ and the independence assumption breaks down.

Individual-based mean-field approximations can be extended by using pair-approximation approaches (ben-Avraham and Köhler, 1992), in which the expectations $E[X_i X_j]$ are considered as relevant dynamical quantities, for which the evolution equations are written. In order to provide these equations in closed form, the three-point correlation functions $E[X_i X_j X_m]$ are factorized as a function of the single and two points correlation functions. By the same token it is possible to derive exact equations for the correlation functions

up to n points (Van Mieghem, 2014a). An approximation is, however, always required to close the set of equations by expressing $(n+1)$-point correlations as functions of correlations of lower order. As the order n grows, these approximations are characterized in general by increasing levels of accuracy.

Although the IBMF method can be generalized to time-dependent adjacency matrices and adaptive models, explicit solutions have been obtained mainly for the SIS models on static networks.

B. Degree-based mean-field approach

Degree-based mean-field (DBMF) theory was the first theoretical approach proposed for the analysis of general dynamical processes on complex networks, and its popularity is due to its applicability to a wide range of dynamical processes on networks (Barrat, Barthélemy, and Vespignani, 2008; Dorogovtsev, Goltsev, and Mendes, 2008). The DBMF approximation for dynamical processes on networks starts with the assumption that all nodes of degree k are statistically equivalent. This assumption implies that, instead of working with quantities Φ_i specifying the state of vertex i (as in the IBMF theory), the relevant variables Φ_k are specifying the state of all vertices with degree k, the degree class k (Pastor-Satorras and Vespignani, 2001b; Boguñá and Pastor-Satorras, 2002). The assumption also implies that any given vertex of degree k is connected with the same probability $P(k'|k)$ to any node of degree k'. The approach is a convenient complexity reduction technique that consists of a drastic reduction in the number of degrees of freedom of the system.

DBMF theory for epidemic models focuses on the partial densities of individuals of degree k in the compartment α, $\rho_k^\alpha(t)$, or, in other words, the probability that an individual in the population with degree k is in the compartment α. These variables are not independent, but fulfill the condition $\sum_\alpha \rho_k^\alpha(t) = 1$. The total fraction of individuals in the compartment α is $\rho^\alpha(t) = \sum_k P(k)\rho_k^\alpha(t)$. The explicit rate equations for the quantities $\rho_k^\alpha(t)$ are obtained by using the law of mass action and assuming the independence of the expectation values (see Sec. II.B).

The DBMF theory implicitly contains an approximation that is not always clearly stated. The statistical equivalence within degree classes considers the network itself in a mean-field perspective, in which the adjacency matrix a_{ij} is completely destroyed, only the degree and the two-vertex correlations of each node being preserved. This is equivalent to replacing the adjacency matrix in the IBMF theory by its ensemble average $\bar{a}_{ij}$, expressing the probability that vertices i and j are connected (annealed network approximation), taking the form (Dorogovtsev, Goltsev, and Mendes, 2008; Boguñá, Castellano, and Pastor-Satorras, 2009)

$$\bar{a}_{ij} = \frac{k_j P(k_i|k_j)}{N P(k_i)}. \tag{13}$$

In the case of uncorrelated networks, the simple form $\bar{a}_{ij} = k_i k_j/(N\langle k\rangle)$ is obtained.

The solutions obtained from DBMF theories depend in general on the statistical topological properties of the underlying networks, and in the case of uncorrelated networks, on the moments of its degree distribution. Although the DBMF theory is a strong approximation in the case of dynamical processes occurring on static networks, it appears to be a suitable approximation to capture the behavior of epidemics mediated by interaction patterns changing on a time scale much faster than the time scales of the spreading process. In this limit, we can consider the epidemic process to spread on a network that is constantly rewired, while preserving the given functional form for $P(k)$ and $P(k'|k)$. This process amounts to a contagion process spreading on an effective mean-field network specified by the annealed network approximation. Furthermore, the DBMF provides a good description of a wide range of dynamical processes that include complex compartment transitions, multiple occupancy of nodes, and time-varying connectivity patterns.

C. Generating function approach

For the SIR model and similar models without steady state, the long time (static) properties of the epidemic outbreak can be mapped into a suitable bond percolation problem (see Sec. II.C). In this framework, the probability p that a link exists is related to the probability of transmission of the disease from an infected node to a connected susceptible one.

The problem of percolation in networks (Molloy and Reed, 1995; Callaway *et al.*, 2000; Cohen *et al.*, 2000) can be tackled with generating functions (Wilf, 2006). We consider the case of bond percolation, in which edges in a network are removed with probability $1-p$ and kept with probability p (see Sec. II.C). We define u as the probability that a randomly chosen edge does not lead to a vertex connected to the (possibly existing) giant component. A randomly chosen edge is not connected to the giant component if either it has been removed or it leads to a vertex of degree k, whose remaining $k-1$ edges either do not exist or do not lead to the giant component, i.e.,

$$u = 1 - p + \sum_k \frac{kP(k)}{\langle k\rangle}(1-p+pu)^{k-1}. \tag{14}$$

This equation is valid for degree uncorrelated networks which have no loops,[4] in which a randomly chosen edge points to a vertex of degree k with probability $kP(k)/\langle k\rangle$; see Sec. III.B.3. The probability $1-P_G$ that a randomly chosen vertex does not belong to the giant component is proportional to the probability that it has degree k, and all of its outgoing edges either have been removed or do not lead to the giant component, i.e.,

$$P_G(p) = 1 - \sum_k P(k)(1-p+up)^k. \tag{15}$$

Equations (14) and (15) can be conveniently written in terms of the degree distribution generating function (Wilf, 2006)

[4]The formalism can be extended to degree-correlated networks; see Sec. VII.B.1 and Goltsev, Dorogovtsev, and Mendes (2008).

$G_0(z) = \sum_k P(k) z^k$ and the excess degree generating function $G_1(z) = \sum_k (k+1) P(k+1) z^k / \langle k \rangle$, taking the form

$$u = 1 - p + G_1(1 - p + pu), \tag{16}$$

$$P_G(p) = 1 - G_0(1 - p + pu). \tag{17}$$

The condition for the existence of a giant component translates into the condition for the existence of a nonzero solution of Eq. (16), which is (Callaway *et al.*, 2000)

$$p > p_c = \frac{G_0'(1)}{G_0''(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{18}$$

In the vicinity of the critical point, the expansion of the generating functions around the nonzero solution yields the scaling behavior of the order parameter $P_G(p) \sim (p - p_c)^{\beta_{\text{perc}}}$ with $\beta_{\text{perc}} = 1$ in the case of homogeneous networks. In the case of heterogeneous networks with degree distribution $P(k) \sim k^{-\gamma}$, we surprisingly find that the percolation threshold tends to zero for $\gamma < 3$ in the limit of an infinite network size $N \to \infty$ (Cohen, ben-Avraham, and Havlin, 2002). The critical exponent β_{perc} assumes in this class of networks the following values (Cohen, ben-Avraham, and Havlin, 2002):

$$\beta_{\text{perc}} = \begin{cases} 1/(3-\gamma) & \text{for } \gamma < 3, \\ 1/(\gamma - 3) & \text{for } 3 < \gamma \leq 4, \\ 1 & \text{for } \gamma \geq 4. \end{cases} \tag{19}$$

For the case $\gamma = 3$, a stretched exponential form $P_G(p) \sim e^{1/p}$ is expected, based on the mapping to the SIR model; see Sec. V.B.1.

The above expressions are very general and can be used also to study immunization strategies and other containment measures in the case of SIR-like models. See also Hamilton and Pryadko (2014) and Karrer, Newman, and Zdeborová (2014) for recent further improvements on these results.

V. EPIDEMIC PROCESSES IN HETEROGENEOUS NETWORKS

A. Susceptible-infected-susceptible model

An impressive research effort has been devoted to understanding the effects of complex network topologies on the SIS model. The SIS dynamics involves only two-state variables and may reach a stationary state, making it ideal for the application of several theoretical approaches. For this reason, there are a large number of results concerning the SIS model, obtained with approaches ranging from approximate mean-field theories to exact methods. In the following, we follow a historical perspective that starts with the basic and easily generalizable mean-field approaches and moves then to recent exact results that put our understanding of the SIS model in complex networks on firm theoretical ground.

1. Degree-based mean-field theory

The first approach to the study of the SIS model in complex networks (Pastor-Satorras and Vespignani, 2001b) used a DBMF theory (commonly referred to in the physics literature as the heterogeneous mean-field approach), whose general methodology can be extended to a wealth of dynamical processes in networks (Barrat, Barthélemy, and Vespignani, 2008). In the DBMF approach, the SIS model is described in terms of the probability $\rho_k^I(t)$ that a node of degree k is infected at time t, assuming the statistical equivalence of all nodes of degree k. The SIS dynamical equation for $\rho_k^I(t)$ is derived by applying the law of mass action,

$$\frac{d\rho_k^I(t)}{dt} = -\rho_k^I(t) + \lambda k [1 - \rho_k^I(t)] \sum_{k'} P(k'|k) \rho_{k'}^I(t), \tag{20}$$

where, without loss of generality, we rescaled time by μ^{-1}, so that the recovery rate is unitary and the infection rate is equivalent to the spreading rate $\lambda = \beta/\mu$. The first term accounts for the recovery of nodes of degree k, proportional to the probability $\rho_k^I(t)$ that a node of degree k is infected. The second term accounts for the infection of new nodes and is proportional to the probability that a node of degree k is susceptible, $1 - \rho_k^I(t)$, times the probability $P(k'|k)$ that this node is connected to a node of degree k', multiplied by the probability $\rho_{k'}^I(t)$ that this last node is infected, times the rate of infection λ. This factor is summed over all the possible values of k'. The extra factor k takes into account all the possible edges through which the disease can arrive at a node of degree k.

Equation (20) for the DBMF approximation to the SIS model cannot be solved in a closed form for general degree correlations. The value of the epidemic threshold can however be obtained by means of a linear stability analysis (Boguñá and Pastor-Satorras, 2002). Performing an expansion of Eq. (20) at first order in $\rho_k^I(t)$ leads to

$$\frac{d\rho_k^I(t)}{dt} \simeq \sum_k J_{kk'} \rho_{k'}^I(t), \tag{21}$$

where the Jacobian matrix element is $J_{kk'} = -\delta_{kk'} + \lambda k P(k'|k)$ and δ_{ij} is the Kronecker delta symbol. A null steady state, corresponding to the healthy phase, is stable when the largest eigenvalue of the Jacobian is negative. The endemic phase will thus take place when $-1 + \lambda \Lambda_M > 0$, where Λ_M is the largest eigenvalue of the connectivity matrix (Boguñá and Pastor-Satorras, 2002), whose elements are

$$C_{kk'} = kP(k'|k). \tag{22}$$

From the Perron-Frobenius theorem (Gantmacher, 1974), since C is non-negative, and assuming that it is irreducible, its largest eigenvalue is real and positive. Therefore, the endemic state occurs for

$$\lambda > \lambda_c^{\mathrm{DBMF}} = \frac{1}{\Lambda_M}. \tag{23}$$

In the case of uncorrelated networks, in which $P(k'|k) = k'P(k')/\langle k\rangle$, it is possible to obtain an explicit solution of the DBMF equations by writing

$$\frac{d\rho_k^I(t)}{dt} = -\rho_k^I(t) + \lambda k[1-\rho_k^I(t)]\Theta, \tag{24}$$

where

$$\Theta = \sum_{k'} \frac{k'P(k')}{\langle k\rangle}\rho_{k'}^I(t). \tag{25}$$

The latter expression gives the probability of finding an infected node following a randomly chosen edge. In the steady state, imposing the stationarity condition $d\rho_k^I(t)/dt = 0$, we obtain

$$\rho_k^I = \frac{\lambda k\Theta(\lambda)}{1+\lambda k\Theta(\lambda)}, \tag{26}$$

where Θ is now a constant that depends on the spreading rate λ. Equation (26) shows that the higher the degree of a node, the higher its infection probability, indicating that strongly inhomogeneous connectivity patterns impact the epidemic spreading. The factor $\Theta(\lambda)$ can be computed self-consistently, introducing Eq. (26) into Eq. (25) to obtain

$$\Theta(\lambda) = \frac{1}{\langle k\rangle}\sum_k kP(k)\frac{\lambda k\Theta(\lambda)}{1+\lambda k\Theta(\lambda)}. \tag{27}$$

The self-consistent equation (27) admits a nonzero solution, corresponding to the endemic state, only when the following threshold condition for uncorrelated networks is fulfilled (Pastor-Satorras and Vespignani, 2001b):

$$\lambda > \lambda_c^{\mathrm{DBMF,unc}} = \frac{\langle k\rangle}{\langle k^2\rangle}. \tag{28}$$

The uncorrelated threshold can also be obtained from the general expression (23) by noticing that the elements of the connectivity matrix reduce to $C_{kk'} = k'P(k')/\langle k\rangle$, which has a unique nonzero eigenvector with eigenvalue $\langle k^2\rangle/\langle k\rangle$. For a fully homogeneous (regular) network with $\langle k^2\rangle = \langle k\rangle^2$, Eq. (28) recovers the result $\lambda_c^{\mathrm{DBMF}} = 1/\langle k\rangle$, as expected from the simple arguments from Sec. II.B [see Eq. (11)].

Equation (28) implies that, in networks with a power-law degree distribution with exponent $2 < \gamma \le 3$, for which $\langle k^2\rangle \to \infty$ in the limit of a network of infinite size, the epidemic threshold tends asymptotically to zero. This was one of the first results pointing out the crucial effect of degree heterogeneities on epidemic spreading. The critical behavior of the prevalence in the vicinity of the epidemic threshold can be obtained by solving Eq. (27) for Θ in the continuous degree approximation and introducing the result into the definition $\rho^I(\lambda) = \sum_k P(k)\rho_k^I$. From these manipulations, one obtains (Pastor-Satorras and Vespignani, 2001a) $\rho^I(\lambda) \sim (\lambda - \lambda_c^{\mathrm{DBMF}})^{\beta_{\mathrm{SIS}}^{\mathrm{DBMF}}}$, with the critical exponent

$$\beta_{\mathrm{SIS}}^{\mathrm{DBMF}} = \begin{cases} 1/(3-\gamma) & \text{for } \gamma < 3, \\ 1/(\gamma-3) & \text{for } 3 < \gamma \le 4, \\ 1 & \text{for } \gamma \ge 4. \end{cases} \tag{29}$$

For the case $\gamma = 3$, a prevalence following a stretched exponential form is obtained, namely, $\rho^I(\lambda) \sim e^{-1/(m\lambda)}$ (Pastor-Satorras and Vespignani, 2001b). Noticeably, these exponents take the exact same form as those observed for the percolation problem, Eq. (19). It is interesting to note that for $2 < \gamma \le 3$ the exponent governing the prevalence behavior close to the threshold is larger than 1. As noted by Pastor-Satorras and Vespignani (2001b) this implies that, while the vanishing threshold makes the spreading of pathogens more easy, the very slow growth of the epidemic activity for increasing spreading rates makes epidemic in these networks less threatening.

2. Individual-based mean-field theory

As discussed in Sec. IV, the state of the system in the SIS model is fully defined by a set of Bernoulli random variables $X_i(t) \in \{0,1\}$: $X_i(t) = 0$ for a healthy, susceptible node and $X_i(t) = 1$ for an infected node. It is possible to construct a 2^N Markov chain (Van Mieghem, Omic, and Kooij, 2009; Simon, Taylor, and Kiss, 2011; Van Mieghem and Cator, 2012), exactly specifying the time evolution of the SIS model. While exact, as mentioned, the Markov chain approach complicates analytical calculations. A simpler route to derive rigorous results on the SIS model is to use the property of a Bernoulli random variable X_i that the expectation $E[X_i]$ is equal to the probability that node i is infected, i.e., $E[X_i] = \Pr[X_i = 1] \equiv \rho_i^I(t)$. This allows one to write the exact equations for the expectation of being infected for each node i of the SIS model (Van Mieghem, 2014a, 2014b),

$$\frac{dE[X_i(t)]}{dt} = E\left[-\mu X_i(t) + [1-X_i(t)]\beta\sum_{j=1}^N a_{ij}X_j(t)\right]. \tag{30}$$

Equation (30) holds also for asymmetric adjacency matrices, i.e., for both directed and undirected networks and for time-varying networks where the adjacency matrix $A(t)$ depends on time t (Guo *et al.*, 2013). The SIS governing equation (30) states that the change over time of the probability of infection $E[X_i(t)] = \Pr[X_i(t) = 1]$ of node i equals the average of two competing random variables: (a) if the node i is infected ($X_i = 1$), then $dE[X_i]/dt$ decreases with a rate equal to the curing rate μ, and (b) if the node is healthy ($X_i = 0$), it can be infected with infection rate β from each infected neighbor. The total number of infected neighbors of node i is $\sum_{j=1}^N a_{ij}X_j$.

For a static network, Eq. (30) reduces to (Sharkey, 2011; Schwartz and Stone, 2013; Van Mieghem, 2014b)

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda\sum_{j=1}^N a_{ij}\rho_j^I(t) - \lambda\sum_{j=1}^N a_{ij}E[X_i(t)X_j(t)], \tag{31}$$

where t has been rescaled by $1/\mu$ and $\lambda = \beta/\mu$.

The above equations do not lend themselves to an explicit solution because the equation for $\rho_i^I(t)$ depends on the two-node expectation $E[X_i(t)X_j(t)]$. Its exact computation requires the knowledge of the joint probability distribution $\Pr[X_i = 1, X_j = 1]$ for the state of nodes i and j. In order to derive a closed set of N dynamical equations, the IBMF approximation is usually made [also termed quenched mean-field (QMF) or N-intertwined mean-field approximation (NIMFA)], which assumes that the states of neighboring nodes are statistically independent, i.e.,

$$E[X_i(t)X_j(t)] \equiv E[X_i(t)]E[X_j(t)] = \rho_i^I(t)\rho_j^I(t). \tag{32}$$

Under this approximation the dynamical equation (31) for the SIS model becomes (Hethcote and Yorke, 1984; Wang *et al.*, 2003; Chakrabarti *et al.*, 2008; Van Mieghem, Omic, and Kooij, 2009)

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda[1 - \rho_i^I(t)]\sum_{j=1}^{N} a_{ij}\rho_j^I(t). \tag{33}$$

The physical interpretation is immediate: the change in the probability ρ_i^I has a destruction term, equal to the probability that node i is infected times the rate of recovery $\mu = 1$, and a creation term, equal to the probability that node i is susceptible, times the total probability that any of its nearest neighbors is infected, times the effective transmission rate $\lambda = \beta/\mu$. Again, time has been rescaled in Eq. (33). Noticeably, Eq. (33) can be derived using other approaches. For example, Gómez *et al.* (2010) proposed a discrete-time equation additionally taking into account the possibility of reinfection in a single time step of length Δt. The equation thus obtained leads to Eq. (33) in the continuous-time limit $\Delta t \to 0$.

To obtain a prediction of the threshold, we can apply a linear stability analysis on Eq. (33). Indeed, linearizing Eq. (33) leads to the Jacobian matrix with elements $J_{ij} = -\delta_{ij} + \lambda a_{ij}$. An endemic state occurs when the largest eigenvalue of J is positive. This condition translates in the epidemic threshold

$$\lambda \geq \lambda_c^{\mathrm{IBMF}}, \qquad \lambda_c^{\mathrm{IBMF}} = \frac{1}{\Lambda_1}, \tag{34}$$

where Λ_1 is the largest eigenvalue of the adjacency matrix (Wang *et al.*, 2003; Chakrabarti *et al.*, 2008; Van Mieghem, Omic, and Kooij, 2009).

In networks with a power-law degree distribution $P(k) \sim k^{-\gamma}$, Eq. (34) can be combined with $\Lambda_1 \sim \max\{\sqrt{k_{\max}}, \langle k^2\rangle/\langle k\rangle\}$ (Chung, Lu, and Vu, 2003), where $k_{\max}$ is the maximum degree in the network, to produce an expression for the scaling of the epidemic threshold (Castellano and Pastor-Satorras, 2010, 2012)

$$\lambda_c^{\mathrm{IBMF}} \simeq \begin{cases} 1/\sqrt{k_{\max}} & \gamma > 5/2, \\ \langle k\rangle/\langle k^2\rangle & 2 < \gamma < 5/2. \end{cases} \tag{35}$$

The relevance of this result is the prediction, in the thermodynamic limit, of a vanishing epidemic threshold for every network for which the maximum degree is a growing function of the network size, which is essentially the case for all random, nonregular networks. Although the expression for the epidemic threshold obtained from the IBMF theory is not exact [see Givan *et al.* (2011) for a detailed assessment of the independence assumption], it provides a relatively good accuracy when compared with the results of extensive numerical simulations; see Sec. V.A.5.

It is worth bridging the IBMF approach with the DBMF approach presented in the previous section. As stated in Sec. IV, the DBMF approach is based on the assumption of the statistical equivalence of all nodes with the same degree k, actually defining the spreading process on an effective mean-field graph, whose adjacency matrix is given by the annealed form $\bar{a}_{ij} = k_j P(k_i|k_j)/(NP(k_i))$. This elucidates the connection between the IBMF and DBMF approaches. The latter can be simply derived by substituting the annealed adjacency matrix in Eq. (33). By performing a degree-based average $\rho_k^I = \sum_{i\in k}\rho_i^I/(NP(k))$, Eq. (20) is thus recovered from the IBMF approach. Hence, DBMF is equivalent to IBMF with the additional approximation that the detailed topological network structure is replaced by its annealed version.

Within the framework of IBMF theory, it is also possible to derive the behavior of the prevalence ρ^I in the stationary state just above the epidemic threshold (Goltsev *et al.*, 2012; Van Mieghem, 2012a)

$$\rho^I(\lambda) \simeq \frac{1}{N}\frac{\sum_{j=1}^{N}(x_1)_j}{\sum_{j=1}^{N}(x_1)_j^3}\frac{\lambda - \lambda_c}{\lambda_c}, \tag{36}$$

where $\vec{x}_1$ is the principal eigenvector (PEV) corresponding to the largest eigenvalue of the adjacency matrix. The complete expansion of the prevalence in the stationary state around the epidemic threshold is derived by Van Mieghem (2012b).

Based on Eq. (36), the validity of the IBMF prediction for the epidemic threshold was recently questioned (Goltsev *et al.*, 2012) according to the following argument. For $\lambda_c^{\mathrm{IBMF}}$ to be the true epidemic threshold, the stationary state above it must be endemic with a finite fraction of the network infected. This requires that for $N \to \infty$ the prefactor

$$\mathcal{A} = \frac{1}{N}\frac{\sum_{j=1}^{N}(x_1)_j}{\sum_{j=1}^{N}(x_1)_j^3} \tag{37}$$

in Eq. (36) must tend to a constant of $\mathcal{O}(1)$. Whether $\mathcal{A}$ is constant or not depends on the localization of the PEV, i.e., whether its weight is evenly distributed (delocalized) on all nodes of the network, or localized in a few nodes. Goltsev *et al.* applied this idea to the analysis of power-law distributed networks, arguing by means of analytical calculations and numerical experiments [see also Martin, Zhang, and Newman (2014)] that, for $\gamma \leq 5/2$, the PEV is delocalized, while it is localized for $\gamma > 5/2$. This implies that, while $\lambda_c^{\mathrm{IBMF}}$ always marks a transition to an active state, this one is endemic only for $\gamma < 5/2$, corresponding to a delocalized PEV; for $\gamma > 5/2$, instead, a localized PEV indicates that the transition at $\lambda_c^{\mathrm{IBMF}}$ is not to an endemic state, but to a subendemic state, in which activity is restricted to the neighborhood of the hubs with largest

degree. Support of this argument [which is mean field in nature, based on Eq. (36)] is provided by Lee, Shim, and Noh (2013), who characterize the subendemic state as a Griffiths phase [see also Boguñá, Castellano, and Pastor-Satorras (2013)].

3. Extensions of degree-based and individual-based mean-field approaches

Several extensions of the degree-based and individual-based mean-field theories have been proposed, taking into account the role of dynamical correlations, which are neglected in both approaches.

A natural way to include the effect of correlations is to consider additional variables representing the state of pairs, triples, etc., of neighboring nodes. Eames and Keeling (2002) introduced an extended degree-based approach where the evolution of the average number $\langle I^k \rangle$ of nodes of degree k in the infected state depends on the number $\langle S^k I^l \rangle$ of connections between susceptibles of degree k with infected nodes of degree l. The dynamics can be written in terms of the properties of triples, such as $\langle S^k S^l I^m \rangle$ and so on and so forth. If averages for triples are approximated with averages for pairs and single nodes, the dynamical equations are reduced to a set of $O(k_{\max}^2)$ nonlinear ordinary differential equations (ODEs). This procedure can be iterated, but the increased accuracy is counteracted by a rapid growth in the number of equations.

Similarly, Gleeson (2011), building on the results of Marceau *et al.* (2010), proposed a general theory for binary-state dynamics in networks. This approach takes into account explicitly the dynamical correlations between adjacent nodes [see also Lindquist *et al.* (2011) for a similar approach]. The theory is based on a set of master equations for the quantities $s_{k,m}(t)$ and $i_{k,m}(t)$ which, in the context of the SIS model, are defined as the fraction of nodes of degree k which are susceptible (respectively, infected) at time t and are connected to $m \leq k$ infected neighbors. By means of combinatorial arguments, these quantities can be related to the prevalence ρ_k^I of nodes of degree k, allowing the determination of the prevalence and epidemic threshold. This theoretical approach provides a good description of the time evolution of the prevalence (Gleeson, 2013) and good estimates of the epidemic threshold for random regular lattices (Gleeson, 2011). Gleeson's approach presents again the drawback that the estimation of the threshold in more complex networks requires the numerical solution of large sets of coupled equations, which hinders the analysis of large network sizes.

Another degree-based approach, proposed by Boguñá, Castellano, and Pastor-Satorras (2013), takes into account long distance correlations by explicitly considering the possibility of reinfection between nodes i and j, separated by a topological distance ℓ_{ij} possibly larger than 1. For this purpose, the original SIS dynamics is replaced by a modified description valid over coarse-grained time scales. In such longer temporal intervals, a given infected node i can propagate the infection to any other node j at distance ℓ_{ij} in the network, via a sequence of microscopic infection events of intermediate, nearest neighbor nodes. The infection rate β is then replaced by an effective rate $\bar{\beta}(\ell_{ij}, \beta)$. On the coarse-grained time scale the recovery rate μ of node i is also replaced by an effective rate $\bar{\mu}(k_i, \beta)$. Both parameters $\bar{\beta}(\ell_{ij}, \beta)$ and $\bar{\mu}(k_i, \beta)$ can be estimated from the properties of the network and the SIS model. Writing down a mean-field theory for such extension of the SIS model, upper bounds for the epidemic threshold λ_c of the original SIS model are deduced, which are in good agreement with numerical simulations; see Sec. V.A.5.

For individual-based approaches, the consideration of dynamical correlations can be introduced in a systematic way, by the analog of a cluster expansion (ben-Avraham and Köhler, 1992). The exact SIS equation (31) is, as discussed previously, not closed, due to the presence of the term involving dynamical correlations between pairs of adjacent nodes. One way to proceed consists of complementing Eq. (31) with an equation for the evolution of the pair correlations $E[X_i(t)X_k(t)]$. The $\binom{N}{2}$ governing equations for $dE[X_iX_j]/dt$ for $i \neq j$ take the form (Cator and Van Mieghem, 2012)

$$\begin{aligned}\frac{dE[X_iX_j]}{dt} = &-2\mu E[X_iX_j] + \beta \sum_{k=1}^{N} a_{ik} E[X_jX_k] \\ &+ \beta \sum_{k=1}^{N} a_{jk} E[X_iX_k] \\ &- \beta \sum_{k=1}^{N} (a_{ik} + a_{jk}) E[X_iX_jX_k], \end{aligned} \tag{38}$$

while for $i = j$ Eq. (30) holds. Equations (30) and (38) are still an exact description of the dynamics involving now the terms $E[X_iX_jX_k]$, that in turn need to be determined, via $\binom{N}{3}$ differential equations involving joint fourth order expectations and so on. In summary, the approach leads to a set of $\sum_{k=1}^{N} \binom{N}{k} = 2^N - 1$ exact equations describing the evolution of the SIS process (to be complemented with the conservation of probability) that form a hierarchy: the equations for the evolution of correlations of order n depending on those of order $n+1$.

To allow computations in practice, this hierarchy must be limited to some small n by imposing a closure condition for the set of equations. The simplest closure condition $E[X_iX_j] = E[X_i]E[X_j]$ leads to the IBMF approximation. Higher order closures include dynamical correlations in a more detailed way, thus providing a more accurate description of the system dynamics. The assumption of different closure relations leads to different degrees of tractability of the ensuing equations. Some of those can be proved to be exact for simple networks (Kiss *et al.*, 2015). For example, focusing on general closure forms, Cator and Van Mieghem (2012) proposed the expression $E[X_iX_jX_k] = E[X_iX_j]E[X_k]$. Analogously, Mata and Ferreira (2013), applying standard techniques from pair approximations in statistical physics, proposed the closure

$$E[X_iX_jX_k] = \frac{E[X_iX_j]E[X_jX_k]}{E[X_j]}. \tag{39}$$

The particular interest of the closure (39) is that it allows one to derive an explicit expression for the epidemic threshold in

terms of the largest eigenvalue of the new Jacobian matrix of the dynamical equations (Mata and Ferreira, 2013):

$$J_{ij} = -\left(1 + \frac{\lambda^2 k_i}{2\lambda + 2}\right)\delta_{ij} + \frac{\lambda(2+\lambda)}{2\lambda+2} a_{ij}. \tag{40}$$

A completely different approach to determine the epidemic threshold for the SIS model was proposed by Parshani, Carmi, and Havlin (2010). The idea is to map the SIS dynamics with fixed infection time, to a percolation process, mirroring the approach successfully used for the SIR model (see Sec. V.B.4). In SIS dynamics, however, the mapping is approximate and one has to take into account the reinfection probability π, i.e., the probability that an infected node reinfects the node from which it originally received the disease. By estimating π and using it in a modified percolation approach, values of the epidemic threshold are derived, in good agreement with numerical simulations, also for heavy-tailed degree distributions.

4. Exact results

Although the mean-field approaches provide a general theoretical picture of the behavior of the SIS model in networks, a few exact results exist that provide rigorous bounds for the threshold and the dynamical behavior of the model. A first exact result concerning the lower bound of the epidemic threshold (Van Mieghem and van de Bovenkamp, 2013) can be achieved by revisiting Eq. (31). Since $0 \le \sum_{k=1}^{N} a_{ki} X_i(t) X_k(t)$, it is possible to write the inequality:

$$\frac{d\rho_i^I(t)}{dt} \le -\rho_i^I(t) + \lambda \sum_{k=1}^{N} a_{ki}\rho_k^I(t). \tag{41}$$

Denoting the vector $W = (\rho_1^I, \rho_2^I, \ldots, \rho_N^I)$, the solution of the inequalities (41) is

$$W(t) \le e^{(\lambda A - I)t} W(0). \tag{42}$$

The exponential factor is dominated by the fastest growing mode, which is $\lambda\Lambda_1 - 1$, where Λ_1 is the largest eigenvalue of the non-negative matrix A, which is real and positive, by the Perron-Frobenius theorem (Gantmacher, 1974). When $\lambda\Lambda_1 - 1 \le 0$, then $W_i = \rho_i^I(t)$ decreases exponentially in t toward zero and the epidemic dies out fast, so that

$$\lambda_c \ge \frac{1}{\Lambda_1}. \tag{43}$$

Interestingly, this lower bound coincides with the IBMF result.

Ganesh, Massoulie, and Towsley (2005) proved that the average time $E[T]$ for the SIS Markov process to hit the absorbing state, when the effective infection rate $\lambda < 1/\Lambda_1$, obeys

$$E[T] \le \frac{\log N + 1}{1 - \lambda\Lambda_1} \tag{44}$$

from which Eq. (43) is deduced.

Above the epidemic threshold instead, the activity must be endemic, so that the average time to absorption is $E[T] = O(e^{cN})$ for some constant $c > 0$. Chatterjee and Durrett (2009) proved that in graphs with power-law degree distribution $E[T] > O(e^{bN^{1-\delta}})$ for any $\delta > 0$. This result pointed to a vanishing threshold in the large N limit, but still left the possibility open for nonendemic long-lived metastable states, such as those predicted by Goltsev *et al.* (2012) and Lee, Shim, and Noh (2013). This possibility was recently ruled out by the work of Mountford *et al.* (2013), showing that for any $\lambda > 0$ and large N, the time to absorption on a power-law graph grows exponentially in N, implying that there is endemic activity for any $\lambda > 0$.

For the complete graph, the exact average survival time was determined using the Markov theory (Van Mieghem, Sahneh, and Scoglio, 2014). In particular, the average survival time for all λ and N is

$$E[T] = \sum_{j=1}^{N} \sum_{r=0}^{j-1} \frac{(N-j+r)!}{j(N-j)!} \lambda^r \tag{45}$$

whose asymptotic behavior for large N is

$$E[T] \sim \frac{1}{\mu} \frac{\frac{\lambda}{\lambda_c}\sqrt{2\pi}}{(\frac{\lambda}{\lambda_c} - 1)^2} \frac{\exp\left(N\{\log\frac{\lambda}{\lambda_c} + \frac{\lambda_c}{\lambda} - 1\}\right)}{\sqrt{N}}$$

for an effective infection rate $\lambda = \beta/\mu$ above the epidemic threshold λ_c. Since an infection can survive the longest in the complete graph, the maximum average lifetime (or survival) time of an SIS epidemic in any network with N nodes is not larger than Eq. (45), or larger than $E[T] = O(e^{N\ln(\lambda/\lambda_c)})$.

For power-law graphs, Chatterjee and Durrett (2009) provided exact bounds for the exponent β_{SIS} governing the singular behavior $\rho^I \sim \lambda^{\beta_{\mathrm{SIS}}}$ of the activity at the transition, namely, $\gamma - 1 \le \beta_{\mathrm{SIS}} \le 2\gamma - 3$. This implies that the mean-field value $\beta_{\mathrm{SIS}} = 1$ does not hold for any $\gamma > 2$, as well as the failure of the DBMF prediction, Eq. (29).

For a few special classes of simple graphs such as the complete graph and the star, the 2^N-state Markov chain can be reduced to a much smaller number of states, enabling an exact solution (Van Mieghem and Cator, 2012; Cator and Van Mieghem, 2013; Schwartz and Stone, 2013; Van Mieghem, 2013). More results can be classified as asymptotic exact results, where the network size $N \to \infty$. An overview of asymptotic exact results is given by Durrett (2010).

5. Numerical simulations of the SIS model on networks

As presented above, the different approximations of the SIS process on networks yield different results for the numerical value of the epidemic threshold. This is particularly important in the case of networks with a heavy-tailed degree distribution $P(k) \sim k^{-\gamma}$, where the two main approximations, IBMF and DBMF, lead to the same result for $\gamma < 5/2$, but to noticeable differences for $\gamma > 5/2$, especially in the case $\gamma > 3$. In this

region, while DBMF predicts a finite threshold, IBMF indicates a vanishing one, albeit at a relatively small rate with the system size.

Computational efforts have been mostly devoted to the numerical determination of the epidemic threshold of the SIS model on power-law distributed networks, in order to assess the validity of the different theoretical approaches. For a detailed study on graphs of small size see C. Li, van de Bovenkamp, and Van Mieghem (2012).

The standard numerical procedure to study absorbing phase transitions, such as the epidemic transition of SIS, is based on the determination of the average of the order parameter (in this case the density of infected nodes), restricted only to surviving runs (Marro and Dickman, 1999), i.e., runs which have not reached the absorbing state up to a given time t. Such a technique is not efficient, because close to the threshold long time surviving configurations are rare and an exceedingly large number of realizations of the process are needed in order to get substantial statistics. This problem is particularly severe for a large network size, for which very large simulation times are required, due to the presence of a long initial transient. These issues make the standard procedure impractical and have not led to reliable conclusions until recently.

In order to overcome the restrictions of the surviving runs method, Ferreira, Castellano, and Pastor-Satorras (2012) and Mata and Ferreira (2013) use the quasistationary state (QS) method (de Oliveira and Dickman, 2005; Ferreira *et al.*, 2011), based on the idea of constraining the system in an active state. This procedure is implemented by replacing the absorbing state, every time the system tries to visit it, with an active configuration randomly taken from the history of the simulation [see also Van Mieghem and Cator (2012) for an implementation of the same idea by means of an external field]. With this technique, the threshold is estimated by studying the susceptibility (Ferreira, Castellano, and Pastor-Satorras, 2012), defined as

$$\chi = N\frac{\langle \rho^{I2}\rangle - \langle \rho^{I}\rangle^2}{\langle \rho^{I}\rangle}. \tag{46}$$

When plotted as a function of λ in a system of size N, the susceptibility χ exhibits a maximum at a value $\lambda_p(N)$, corresponding to a transition rounded by finite-size effects. In the thermodynamic limit, the position of the peak tends to the critical point as $\lambda_p(N) - \lambda_c(\infty) \sim N^{-1/\bar{\nu}}$ (Binder and Heermann, 2010). Large-scale simulations performed using the QS method (Ferreira, Castellano, and Pastor-Satorras, 2012; Mata and Ferreira, 2013), see Fig. 5, show that, for $\gamma < 5/2$, the IBMF and a pair approximation at the individual level (PQMF) are almost exact, coinciding asymptotically with the DBMF result in this range of degree exponents. For $5/2 < \gamma < 3$, on the other hand, the IBMF result provides the correct scaling of the threshold with network size. For the crucial case $\gamma > 3$, where IBMF and DBMF provide radically different predictions, the results are not as conclusive. A new numerical approach has been proposed to explore this region (Boguñá, Castellano, and Pastor-Satorras, 2013), based on the study of the lifetime of individual realizations of the SIS process starting with a single infected node. Each realization is

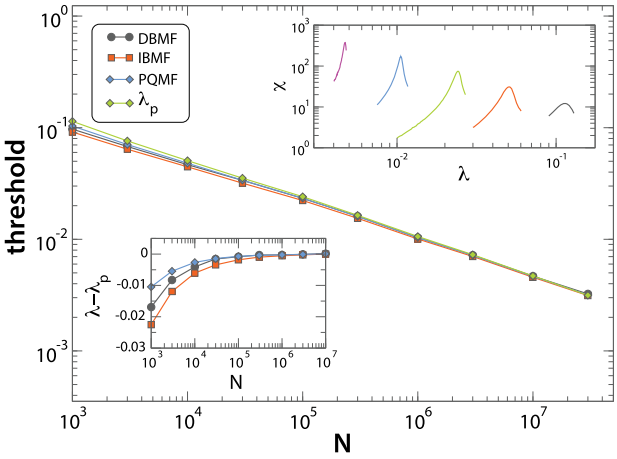

FIG. 5 (color online). Numerical thresholds for the SIS model as a function of the network size N in scale-free networks with degree exponent $\gamma = 2.25$, computed using the QS method, compared with different theoretical predictions. The upper inset shows the behavior of the susceptibility as a function of the spreading rate for different values of $N = 10^3, 10^4, 10^5, 10^6, 10^7$, from right to left. The lower inset shows the difference between the different theoretical thresholds and the peaks of the susceptibility. Adapted from Mata and Ferreira, 2013.

characterized by duration T and coverage C, where the latter is the fraction of distinct nodes infected at least once during the realization. In the thermodynamic limit, realizations can be either finite (i.e., having a finite lifetime and, therefore, vanishing coverage) or endemic (i.e., having an infinite lifetime and coverage equal to 1). The average lifetime $E[T]$ of finite realizations plays the role of a susceptibility, exhibiting a peak at the transition, whose position can then be used to estimate the threshold. The nontrivial problem to

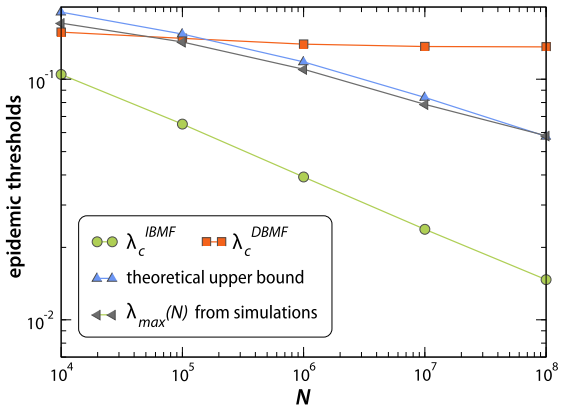

FIG. 6 (color online). Numerical thresholds for the SIS model as a function of the network size N in power-law distributed networks with degree exponent $\gamma = 3.5$, computed using the average lifetime method proposed by Boguñá, Castellano, and Pastor-Satorras (2013). Numerical data are compared with different theoretical approaches as well as with the upper bound obtained from the DBMF theory with long-range dynamical correlations, developed by Boguñá, Castellano, and Pastor-Satorras (2013). Adapted from Boguñá, Castellano, and Pastor-Satorras, 2013.

determine whether, in a finite system, a realization is endemic or not, can be overcome by declaring endemic all realizations for which the coverage reaches a predefined value (e.g., $C = 0.5$). Numerical simulations performed with this method indicate that the extended DBMF approach by Boguñá, Castellano, and Pastor-Satorras (2013) provides a good fit to the numerical threshold for $\gamma > 3$, see Fig. 6, with a scaling with network size that is essentially given by the IBMF expression (35).

6. Finite-size effects and the epidemic threshold

As seen in the previous sections, the connectivity pattern of the network explicitly enters in the determination of the epidemic threshold that generally depends on the moments of the degree distribution and/or the maximum degree of the network. This finding has particular relevance in networks with heavy-tailed degree distributions, where the probability of nodes with very large degree is appreciable. In the limit of infinite-size networks, the epidemic threshold may be vanishing, thus prompting the disruption of the classical epidemic framework where the disease can spread only for adequate transmissibility of the pathogen. While mathematically compelling, the argument of a vanishing threshold was soon recognized as not realistic in real-world networks (May and Lloyd, 2001; Pastor-Satorras and Vespignani, 2002a). Even if the connectivity pattern of a network is well approximated by a heavy-tailed distribution in a given range of degree values, any real-world network is composed by a finite number of nodes N. For instance, the finite size of scale-free networks is generally related to the presence of a natural maximum degree $k_{\max} \sim N^{1/(\gamma-1)}$, as reported in Sec. III.D, that translates into a finite effective epidemic threshold. Although the finite size of the network is often a determinant element in the estimation of the epidemic threshold, for instance in the analysis of numerical simulations (see Sec. V.A.5), there are many other limitations to the maximum degree of the network. These limits are often imposed by spatiotemporal constraints, such as maximum occupancy in spatial locations and the finite time each individual can interact with other individuals. As well, intrinsic cognitive and biological constraints may be at work in real-world systems. One example is provided by the so-called Dunbar's number that limits humans' degree to between 100 and 200 individuals, a size apparently imposed by the finite neocortical processing capacity of the brain (Dunbar, 1998). Interestingly, Dunbar's number has been observed in a wide range of human activities, including communication on modern information technologies, making it a relevant limit in the case of many information diffusion processes (Gonçalves, Perra, and Vespignani, 2011; Miritello *et al.*, 2013).

In view of these inherent limitations, it is often convenient to assume that even in the case of heavy-tailed networks the degree distribution is characterized by the analytic form $P(k) \simeq k^{-\gamma} \exp(-k/k_c)$, where k_c is a characteristic degree size. The exponential cutoff makes it extremely unlikely to observe nodes with degree much larger than k_c, effectively introducing an intrinsic limit to the connectivity capacity of nodes (Pastor-Satorras and Vespignani, 2002a). Within the DBMF approach this leads, for large k_c and $2 < \gamma < 3$, to $\lambda_c^{\mathrm{DBMF,unc}} \simeq (k_c/m)^{\gamma-3}$ where m is the minimum degree of the network, which can be generalized for other values of γ and which shows the effect of the degree limitations imposed by the intrinsic biological, social, and cognitive constraints in real-world networks. Similar finite-size effects and considerations also apply to the epidemic threshold obtained with the IBMF theory and other approaches.

It is important to stress however that the presence of an epidemic threshold because of finite-size effects and other connectivity limitations should not be considered as an argument to neglect the network heterogeneity. It is indeed possible to show with simple calculations (Pastor-Satorras and Vespignani, 2002a) that simple homogenous approaches can overestimate the actual epidemic threshold in heterogeneous networks by one or more orders of magnitude.

B. Susceptible-infected-recovered model

The SIR model is a cornerstone in infectious disease modeling. It applies to the wide range of diseases that provide immunity to the host and it is also a widely used modeling scheme in knowledge and information diffusion (see Sec. X). Theoretically, the SIR model represents a different challenge with respect to the SIS model because it does not allow for a stationary state. The two most used routes to a general analysis of the SIR model have been initially the DBMF theory and the mapping of static properties to the percolation model. Here we start with a presentation of the DBMF approach, focusing then on other degree-based, individual-based, and alternative methods which have been completing the understanding of the SIR dynamics in networks in recent years. We end the section with an overview of the exact results on static properties which can be obtained by mapping SIR to bond percolation.

1. Degree-based mean-field approach

The DBMF approach can be easily adapted to provide insight into the dynamic and static properties of the SIR model. In the DBMF approximation, we define as a function of time three different partial densities, namely, of infected, susceptible, and recovered nodes of degree k, denoted by the variables $\rho_k^I(t)$, $\rho_k^S(t)$, and $\rho_k^R(t)$, respectively.

The order parameter (prevalence) of the SIR model, defined as the number of recovered individuals at the end of the epidemics, is then given by $\rho_\infty^R = \lim_{t\to\infty} \sum_k P(k)\rho_k^R(t)$. In describing the time evolution of these densities, one can follow the analogy with the SIS model, to obtain the set of equations (Lloyd and May, 2001; Moreno, Pastor-Satorras, and Vespignani, 2002)

$$\begin{aligned} \frac{d\rho_k^I(t)}{dt} &= -\rho_t^I(t) + \lambda k \rho_k^S(t)\Gamma_k(t), \\ \frac{d\rho_k^R(t)}{dt} &= \rho_k^I(t), \end{aligned} \tag{47}$$

complemented with the normalization condition $\rho_k^S(t) = 1 - \rho_k^I(t) - \rho_k^R(t)$, where

$$\Gamma_k(t) = \sum_{k'} P(k'|k)\rho_{k'}^I(t). \tag{48}$$

The value of the epidemic threshold in the case of general correlations can be obtained as in the SIS case, by performing a linear stability analysis. The same result follows, with the epidemic threshold given by the inverse of the largest eigenvalue Λ_M of the connectivity matrix, Eq. (22). As for SIS, in the case of uncorrelated networks the epidemic threshold is given by $\lambda_c = \langle k\rangle/\langle k^2\rangle$ (Lloyd and May, 2001; Moreno, Pastor-Satorras, and Vespignani, 2002). For uncorrelated networks, within the same DBMF approximation, it is also possible to integrate the rate equations over time, starting from a small seed, thus obtaining the full temporal evolution of the spreading process. The solution depends on a differential equation for an auxiliary function $\phi(t)$, which cannot be solved analytically in general. However, in the infinite time limit, it is possible to determine the dependence of the final prevalence ρ_∞^R on λ:

$$\rho_\infty^R = \sum_k P(k)(1 - e^{-\lambda k \phi_\infty}), \tag{49}$$

where

$$\phi_\infty = 1 - \frac{1}{\langle k\rangle} - \sum_k \frac{kP(k)}{\langle k\rangle} e^{-\lambda k \phi_\infty}. \tag{50}$$

The solution of Eq. (50) leads again to the epidemic threshold $\lambda_c = \langle k\rangle/\langle k^2\rangle$, a result that recovers the naive expectation for regular networks; see Eq. (11), $\lambda_c = 1/\langle k\rangle$. For a power-law degree distribution $P(k) \sim k^{-\gamma}$, a detailed analysis (Moreno, Pastor-Satorras, and Vespignani, 2002) leads to a prevalence, in the vicinity of the epidemic threshold, of the form $\rho_\infty^R \sim (\lambda - \lambda_c)^{\beta_{\mathrm{SIR}}}$, with exponent β_{SIR} coinciding with the value for bond percolation, Eq. (19). The above results are exact for annealed networks, when the topology changes [preserving $P(k)$] at a very fast rate (Volz and Meyers, 2009). Instead, when considering it as an approach to static networks, the DBMF can be improved taking into account the fact that, in the SIR process, a vertex cannot propagate the disease to the neighbor who originally infected it, because the latter is necessarily not susceptible. This effect can be included in the DBMF equations by discounting, from the number of edges pointing from infected individuals of degree k' to vertices of degree k, the edge from which the original infection arrived to the vertices of degree k'. In this way, Eqs. (47) are recovered but now the $\Gamma_k(t)$ function takes the form (Boguñá, Pastor-Satorras, and Vespignani, 2003b)

$$\Gamma_k(t) = \sum_{k'} \frac{k'-1}{k'} P(k'|k)\rho_{k'}^I(t). \tag{51}$$

The value of the epidemic threshold in this case is given by $\lambda_c = 1/\tilde{\Lambda}_M$, where $\tilde{\Lambda}_M$ is the largest eigenvalue of the new connectivity matrix

$$\tilde{C}_{kk'} = \frac{k(k'-1)}{k'} P(k'|k). \tag{52}$$

In the case of uncorrelated networks, the largest eigenvalue of the matrix $\tilde{C}_{kk'}$ is $\tilde{\Lambda}_M = \langle k^2\rangle/\langle k\rangle - 1$ (the corresponding eigenvector has components $\tilde{v}_k = k$) so that the epidemic threshold is

$$\lambda_c = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}. \tag{53}$$

As shown below, Eq. (53) is an approximation of the exact result (62). However, this modified DBMF approach captures the correct qualitative behavior, discriminating between vanishing threshold, for scale-free networks, and finite threshold, for $\gamma > 3$.

The DBMF approach allows one to also tackle the scaling of the time evolution of the epidemic outbreak. This is particularly important in the context of models like SIR that do not have a stationary state. For simplicity we initially focus on the SI model (Anderson and May, 1992), representing a disease in which infected individuals never recover and keep propagating the disease forever. The SI model can be considered the limit of the SIR model in which the recovery rate μ is set to zero. While this simplification leads to a trivial asymptotic state in which the whole population becomes eventually infected, it is nevertheless interesting due to its simplicity, which allows one to obtain explicit results for the initial time evolution of epidemic outbreaks. The DBMF analysis of the SI model proceeds from the analog of Eq. (47), valid for generic networks (Barthélemy *et al.*, 2004, 2005)

$$\frac{d\rho_k^I(t)}{dt} = \beta k[1 - \rho_k^I(t)]\Gamma_k(t), \tag{54}$$

with

$$\Gamma_k(t) = \sum_{k'} P(k'|k)\rho_{k'}^I(0) + \sum_{k'} \frac{k'-1}{k'} P(k'|k)[\rho_{k'}^I(t) - \rho_{k'}^I(0)]. \tag{55}$$

The first term in Eq. (55) accounts for a very small initial seed of infected individuals, with initial partial density $\rho_k^I(0)$, which can infect all their neighbors. The second term represents the contribution of individuals infected during the outbreak, which can infect all their neighbors, with the exception of those who transmitted the disease. Linear stability analysis shows that the time evolution at very short times (when the partial densities of infected individuals are very small) follows an exponential growth $\rho^I(t) \sim e^{t/\tau}$, where the characteristic time is given by $\tau = (\beta\tilde{\Lambda}_M)^{-1}$, where again $\tilde{\Lambda}_M$ is the largest eigenvalue of the connectivity matrix in Eq. (52). In the case of uncorrelated networks this implies (Barthélemy *et al.*, 2004, 2005)

$$\tau = \frac{\langle k\rangle}{\beta[\langle k^2\rangle - \langle k\rangle]}. \tag{56}$$

The solution for the SI model can be extended to the case of the general SIR model by allowing a nonzero healing rate, which leads to the general time scale of the initial growth (Barthélemy *et al.*, 2005)

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta)\langle k \rangle}. \tag{57}$$

These results readily indicate that the growth time scale of an epidemic outbreak is inversely proportional to the second moment of the degree distribution $\langle k^2 \rangle$; when this quantity diverges, as in the case of scale-free networks, not only does the threshold tend to vanish, but also the time until the establishment of the infection becomes very small (vanishing in the thermodynamic limit). Computer simulations allow one to obtain a detailed picture of the mechanism of spreading of a disease in a scale-free network (Barthélemy *et al.*, 2004, 2005): Initially, the infection reaches the hubs and from them it quickly invades the rest of the network via a cascade through progressively smaller degree classes.

2. Individual and pair-based mean-field approaches

As in the SIS case, a systematic way to attack the SIR model is based on the full master equation for the exact evolution of probabilities of microscopic states, and the derivation, starting from it, of deterministic evolution equations for dynamical quantities. In this framework, Sharkey (2008) considered SIR with Poissonian infection and recovery processes and derived from the master equation the $2N$ equations for the probabilities for the state of individuals

$$\begin{aligned} \frac{d\rho_i^S(t)}{dt} &= -\beta \sum_j a_{ij} \langle S_i I_j \rangle, \\ \frac{d\rho_i^I(t)}{dt} &= \beta \sum_j a_{ij} \langle S_i I_j \rangle - \mu \rho_i^I, \end{aligned} \tag{58}$$

where S_i and I_j are Bernoulli variables equal to 1 when the node is susceptible (infected, respectively) and 0 otherwise, $\rho_i^S = \langle S_i \rangle$ is the probability that node i is in state S, $\rho_i^I = \langle I_i \rangle$ is the analog for state I, and $\langle S_i I_j \rangle$ is the joint probability of state $S_i I_j$. In order to close Eqs. (58), the simplest possibility is to assume that the state of neighbors is independent (individual-based mean-field approximation). Alternatively, one can derive from the master equation the evolution of the probabilities of pairs of neighbors, which depend in turn on the state of triples of neighboring nodes. The closure of the hierarchy at this level (pair-based mean-field) requires the approximation of probabilities for triples with moments of lower order. There are several possible ways to implement the closure and the best choice is not a trivial problem. The validity of the different approximation schemes has been investigated by Sharkey (2011), who showed that replacing $\langle S_i I_j \rangle = \langle S_i \rangle \langle I_j \rangle$ is equivalent to writing down an equation for the evolution of $\langle S_i I_j \rangle$ containing unphysical terms (i.e., terms assuming that a node is at the same time susceptible and infected). The consequences of these unphysical terms are relevant: from the individual-based mean-field approach one can derive an expression for the SIR epidemic threshold equal to what is found for the SIS case (Youssef and Scoglio, 2011; Prakash *et al.*, 2012): $\lambda_c = 1/\Lambda_1$, where Λ_1 is the largest eigenvalue of the adjacency matrix. This result, however, is even qualitatively not correct, as it predicts a vanishing threshold for power-law distributed networks with $\gamma > 3$, at odds with exact results (see Sec. IV.B.4) and numerical simulations (Castellano and Pastor-Satorras, 2010). The pair-based approach instead, complemented with the closure in Eq. (39), is proven to be an exact description of the stochastic system for a tree topology (Sharkey *et al.*, 2015). In the case of networks with loops it is possible to find a precise connection between the detailed loop structure and the closures that leave the description exact (Kiss *et al.*, 2015). From these individual and pair-based approaches, by summing over all nodes, the equations for the probabilities of the global quantities ρ^I and ρ^S can be obtained, thus providing a microscopic foundation of equations obtained at a population level by means of the mass-action principle. Equation (58) and similar pair-based approaches can be written also for heterogeneous infection and recovery rates (Sharkey, 2008). Hence, the approaches apply in full generality also to directed and weighted networks.

3. Other approaches

Because of its great relevance, the time evolution of the SIR dynamics has been tackled with many other approaches. The extended degree-based approach of Eames and Keeling (2002) (see Sec. V.A) can be applied also to the SIR model, providing a set of closed ODEs that can be integrated numerically or used to derive an expression for the basic reproduction ratio R_0. Also the extended degree-based approach of Lindquist *et al.* (2011) can be applied to SIR, by categorizing each node by its disease state (i.e., S, I, R), as well as by the number of neighbors in each disease state. In this way, an excellent agreement with numerical simulations for both the temporal evolution and the final outbreak size is found. The threshold condition derived analytically turns out to be equal to the exact one obtained using percolation theory, Eq. (62).

An alternative approach by Volz (2008) describes the Poissonian SIR epidemics at the global population level. Based on the probability generating function for the degree distribution, it describes the evolution of the infection using only three coupled nonlinear ordinary differential equations. The solution of these equations is in excellent agreement with numerical simulations (Lindquist *et al.*, 2011); it is shown to be exact in the thermodynamic limit (Decreusefond *et al.*, 2012; Janson, Luczak, and Windridge, 2014) and it allows one to derive the exact expression, Eq. (62), for the epidemic threshold, in the case of static uncorrelated networks. In this case, the approach of Volz (2008) can be shown (House and Keeling, 2010) to be a specific case of the extended degree-based theory of Eames and Keeling (2002). Volz's approach can be made more physically transparent and simpler, reducing to a single evolution equation (Miller, 2011). The basic idea of this improved approach is to focus on the state of a random partner instead of a random individual. From this starting point, a fully general theoretical framework (edge-based compartmental modeling) can be developed, allowing one to deal with many different scenarios, including static and dynamic networks, both undirected and directed (Miller, Slim,

and Volz, 2012; Valdez, Macri, and Braunstein, 2012b; Miller and Volz, 2013). For other approaches to SIR dynamics based on the probability generating function, see Marder (2007) and Noël *et al.* (2009, 2012).

A derivation of a condition for the possibility of a global spreading event starting from a single seed in SIR-like models on generic networks is presented by Dodds, Harris, and Payne (2011) and generalized by Payne, Harris, and Dodds (2011). The approach is based on the state of "node-edge" pairs and relates the possibility of spreading to the condition that the largest eigenvalue of a "gain ratio" matrix (encoding information on both the topology and the spreading process) is larger than 1.

Finally, a new, substantial step forward in the understanding of the SIR model is the recent application of the message-passing approach to SIR dynamics (Karrer and Newman, 2010). This approach provides an exact description of the dynamics on trees, via a closed set of integrodifferential equations, allowing the calculation of the probabilities to be in states S, I, or R for any node and any time. When loops are present, the method gives instead a rigorous bound on the size of disease outbreaks. On generic (possibly directed) trees the approach of Karrer and Newman (2010) has been shown (Wilkinson and Sharkey, 2014) to coincide for Poissonian infections with the pair-based moment closure presented by Sharkey *et al.* (2015). Remarkably, the message-passing approach allows one to deal with fully generic (non-Poissonian) infection and recovery processes.

4. Mapping the SIR model to a percolation process

The connection between the static properties of the SIR model and bond percolation (see Sec. IV.C) was recognized long ago (Ludwig, 1975; Grassberger, 1983; Andersson and Britton, 2000). In the context of epidemics on complex networks, the mapping has been studied in detail by Newman (2002b). Considering a SIR model with uniform infection time τ, i.e., where infected nodes become recovered at time τ after infection,[5] and infection rate β, the transmissibility T is defined as the probability that the infection will be transmitted from an infected node to a connected susceptible neighbor before recovery takes place. For continuous-time dynamics the transmissibility can be computed as (Newman, 2002b)

$$T = 1 - \lim_{\delta t \to 0} (1 - \beta \delta t)^{\tau/\delta t} = 1 - e^{-\tau\beta}. \tag{59}$$

The set of recovered nodes generated by an SIR epidemic outbreak originated from a single node is nothing else than the cluster of the bond percolation problem (with occupation probability T) to which the initial node belongs. The correspondence is exact: all late-time static properties of the SIR model can be derived as direct translations of the geometric properties of the percolation problem. For treelike networks the exact epidemic threshold is given by Eq. (18), so that

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \Rightarrow \beta_c = \frac{1}{\tau} \ln \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle}. \tag{60}$$

The behavior of the outbreak size close to the epidemic threshold, ruled by the equivalent percolating giant component, is given in terms of the exponents in Eq. (19). Equation (60) confirms for the SIR model that the epidemic threshold has a qualitatively different behavior for scale-free networks ($\gamma \leq 3$) and for scale-rich ones ($\gamma > 3$). In the former case the second moment of the degree distribution diverges, so that the threshold vanishes: scale-free networks are extremely vulnerable to disease spreading.

The above results can be considered exact only for a tree (completely loopless) structure. In other networks, the presence of loops and multiple spreading paths leads in general to correlations, which may invalidate the results obtained for trees. However, for random networks which are locally treelike the presence of long loops (infinitely long in the thermodynamic limit) is not sufficient to alter the validity of the results obtained using the tree ansatz (Dorogovtsev, Goltsev, and Mendes, 2008). A different conclusion holds instead in networks with short loops (finite clustering) as discussed in Sec. VII.B.2.

The derivation of Eq. (60) is based on a uniform infection time. More realistically, we assume that infection times τ_i and rates β_{ij} vary between individuals. This implies that the transmissibility T_{ij} depends on the specific edge (i, j). One possible approach, that reduces to the solution of the homogeneous case (Newman, 2002b), is to neglect fluctuations, and replace T_{ij} by its mean value

$$\langle T_{ij} \rangle = 1 - \int d\tau \int d\beta e^{-\beta\tau} Q(\beta) P(\tau), \tag{61}$$

where Q and P are the distributions of β_{ij} and τ_i, respectively. The case of nondegenerate τ_i includes the usual definition of the SIR model with constant recovery rate μ for which recovery times are distributed exponentially with average $\langle \tau_i \rangle = 1/\mu$. In such a case, performing the integral in Eq. (61) and setting $\beta \langle \tau_i \rangle = \beta/\mu = \lambda$ yields $\langle T_{ij} \rangle = \lambda/(1 + \lambda)$, implying

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle}. \tag{62}$$

This approximation leads to the exact epidemic threshold, the mean outbreak size below it, and the final size above it, but fails in other respects (Kenah and Robins, 2007); see also Trapman (2007). The discrepancy is due to correlations (Karrer and Newman, 2010): "if an individual recovers quickly, then the probability of transmission of the disease to any of its neighbors is small; if it takes a long time to recover the probability is correspondingly larger." Newman's approximation is not exact also when the τ_i are degenerate and the β_{ij} vary (Miller, 2007).

The correct way to take into account the heterogeneous transmissibility maps the disease spreading to a bond percolation process, involving now a semidirected network (epidemic percolation network) (Kenah and Robins, 2007;

[5]Note that this does not coincide exactly with the definition given in Sec. II.A.

Miller, 2007); see Sec. III.A. The mapping works as follows. For each pair of connected nodes i and j in the contact network, place a directed edge from i to j with probability $1-e^{-\beta_{ij}\tau_i}$ and a directed edge from j to i with probability $1-e^{-\beta_{ji}\tau_j}$. Tools from percolation theory on directed networks (Boguñá and Serrano, 2005), see Sec. VII.B.4, allow one to exactly characterize the long time features of the epidemic process. In particular, the epidemic transition is associated with the formation of a GSCC in the directed network. If such a component exists, then an infection originating in one of its nodes or in the GIN will spread to all nodes in the GSCC and in the GOUT, giving rise to a macroscopic outbreak. It is crucial to recognize that the GIN and GOUT components play completely different roles: nodes in GOUT are necessarily part of macroscopic outbreaks but cannot originate them. The opposite is true for nodes in GIN. As a consequence the probability that an epidemic occurs (given by the size of GIN ∪ GSCC) and the size of the epidemic (equal to the size of GSCC ∪ GOUT) do not coincide (Meyers, Newman, and Pourbohloul, 2006; Miller, 2007). The mapping to percolation on semidirected networks is valid for any type of contact network underlying the SIR epidemics. For trees and locally treelike networks it is again possible to apply the machinery of probability generating functions to derive explicit results for the related percolation properties. Other discrepancies of the mapping to percolation approach to the SIR model are reported by Lagorio *et al.* (2009).

VI. STRATEGIES TO PREVENT OR MAXIMIZE SPREADING

A. Efficient immunization protocols

The fact that epidemic processes in heavy-tailed networks have a vanishing threshold in the thermodynamic limit, or a very small one in large but finite networks (see Sec. V), prompted the study of immunization strategies leveraging on the network structure in order to protect the population from the spread of a disease. Immunization strategies are defined by specific rules for the identification of the individuals that shall be made immune, taking into account (local or nonlocal) information on the network connectivity pattern. Immunized nodes are in practice removed from the network, together with all the links incident to them, and each strategy is assessed by the effects of immunizing a variable fraction g of nodes in the network. The application of immunization does not only protect directly immunized individuals, but can also lead, for a sufficiently large fraction g, to an increase of the epidemic threshold up to an effective value $\lambda_c(g) > \lambda_c(g=0)$, precluding the global propagation of the disease. This effect is called *herd immunity*. The main objective in this context is to determine the new epidemic threshold as a function of the fraction of immunized individuals. Indeed, for a sufficiently large value of g, any strategy for selecting immunized nodes will lead to an increased threshold. We define the immunization threshold $g_c(\lambda)$, for a fixed value of λ such that, for values of $g > g_c(\lambda)$ the average prevalence is zero, while for $g \le g_c(\lambda)$ the average prevalence is finite.

The simplest immunization protocol, using essentially no information at all, is the random immunization, in which a number gN of nodes is randomly chosen and made immune. While random immunization in the SIS model (under the DBMF approximation) can depress the prevalence of the infection, it does so too slowly to increase the epidemic threshold substantially. Indeed, from Eq. (20), an epidemics in a randomly immunized network is equivalent to a standard SIS process in which the spreading rate is rescaled as $\lambda \to \lambda(1-g)$, i.e., multiplied by the probability that a given node is not immunized, so that the immunization threshold becomes (Pastor-Satorras and Vespignani, 2002b)

$$g_c(\lambda) = 1 - \frac{\langle k \rangle}{\lambda \langle k^2 \rangle}. \tag{63}$$

For heterogeneous networks, for which $\langle k^2 \rangle$ diverges and any value of λ, $g_c(\lambda)$ tends to 1 in the limit $N \to \infty$, indicating that almost the whole network must be immunized to suppress the disease.

This example shows that an effective level of protection in heavy-tailed networks must be achieved by means of optimized immunization strategies (Anderson and May, 1992), taking into account the network heterogeneity. Large degree nodes (the hubs leading to the large degree distribution variance) are potentially the largest spreaders. Intuitively, an optimized strategy should be targeting those hubs rather than small degree vertices. Inspired by this observation, the targeted immunization protocol proposed by Pastor-Satorras and Vespignani (2002b) considers the immunization of the gN nodes with largest degree. A simple DBMF analysis leads to an immunization threshold given, for the SIS model, by the implicit equation (Pastor-Satorras and Vespignani, 2002b)

$$\frac{\langle k^2 \rangle_{g_c}}{\langle k \rangle_{g_c}} = \frac{1}{\lambda}, \tag{64}$$

where $\langle k^n \rangle_g$ is the nth moment of the degree distribution $P_g(k)$ of the network resulting after the deletion of the gN nodes of highest degree, which takes the form (Cohen *et al.*, 2001)

$$P_g(k) = \sum_{k' \ge k}^{k_c} P(k') \binom{k'}{k} (1-g)^k g^{k'-k}. \tag{65}$$

Equation (64) can be readily solved in the case of scale-free networks. For a degree exponent $\gamma = 3$, the immunization threshold reads $g_c(\lambda) \simeq \exp[-2/(m\lambda)]$, where m is the minimum degree in the network. This result highlights the convenience of targeted immunization, with an immunization threshold that is exponentially small over a large range of the spreading rate λ. A similar effect can be obtained with a proportional immunization strategy (Pastor-Satorras and Vespignani, 2002b) [see also Dezsö and Barabási (2002) for a similar approach involving the cure of infected individuals with a rate proportional to their degree], in which nodes of degree k are immunized with probability g_k, which is some increasing function of k. In this case, the infection is eradicated when $g_k \ge 1 - 1/(\lambda k)$, leading to an immunization threshold (Pastor-Satorras and Vespignani, 2002b)

$$g_c(\lambda) = \sum_{k>\lambda^{-1}} \left(1 - \frac{1}{k\lambda}\right) P(k), \tag{66}$$

which takes the form $g_c(\lambda) \simeq (m\lambda)^2/3$ for scale-free networks with $\gamma = 3$.

Other approaches to immunization stress that not only the behavior close to the critical point should be taken into account, but also the entire prevalence curve (the so-called viral conductance) (Kooij *et al.*, 2009; Youssef, Kooij, and Scoglio, 2011; Van Mieghem, 2012b). Additionally, strategies involving possible different interventions on different nodes have been analyzed within a game-theoretic formalism (Van Mieghem and Omic, 2008; Omic, Orda, and Van Mieghem, 2009; Gourdin, Omic, and Mieghem, 2011).

The previously discussed immunization protocols are based on a global knowledge of the network properties (the whole degree sequence must be known to selectively target the nodes to be immunized). Actually, the more a global knowledge of the network is available, the more effective is the immunization strategy. For instance, one of the most effective targeted immunization strategies is based on the betweenness centrality (see Sec. III.B.5), which combines the bias toward high degree nodes and the inhibition of the most probable paths for infection transmission (Holme *et al.*, 2002). This approach can even be improved by taking into account the order in which nodes are immunized in a sequential scheme in which the betweenness centrality is recomputed after the removal of every single node, and swapping the order of immunization in different immunization sequences, seeking to minimize a properly defined size for the connected component of susceptible individuals. This approach has been proven to be highly efficient in the case of the SIR model (Schneider *et al.*, 2011). Improved immunization performance in the SIR model has been found with an "equal graph partitioning" strategy (Chen *et al.*, 2008) which seeks to fragment the network into connected components of approximately the same size, a task that can be achieved by a much smaller number of immunized nodes, compared with a targeted immunization scheme.

The information that makes targeted strategies very effective also makes them hardly feasible in real-world situations, where the network structure is only partially known. In order to overcome this drawback, several local immunization strategies have been considered. A most ingenious one is the acquaintance strategy proposed by Cohen, Havlin, and ben-Avraham (2003), and applied to the SIR model. In this protocol, a number gN of individuals is chosen at random and each one is asked to point to one of his or her nearest neighbors. Those nearest neighbors, instead of the nodes, are selected for immunization. Given that a randomly chosen edge points with high probability to a large degree node, this protocol realizes in practice a preferential immunization of the hubs that results in being effective in hampering epidemics. An analogous result can be obtained by means of a random walk immunization strategy (Holme, 2004; Ke and Yi, 2006), in which a random walker diffuses in the network and immunizes every node that it visits, until a given degree of immunization is reached. Given that a random walk visits a node of degree k_i with probability proportional to k_i (Noh and Rieger, 2004), this protocol leads to the same effectiveness as the acquaintance immunization.

The acquaintance immunization protocol can be improved by allowing for the consideration of additional information, always at the local level. For example, allowing for each node to have knowledge of the number of connections of its nearest neighbors, a large efficiency is attained by immunizing the neighboring nodes with the largest degree (Holme, 2004). As more information is available, one can consider the immunization of the nodes with highest degree found within short paths of length ℓ starting from a randomly selected node (Gomez-Gardenes, Echenique, and Moreno, 2006). The random walk immunization strategy, on the other hand, can be improved by allowing a bias favoring the exploration of high degree nodes during the random walk process (Stauffer and Barbosa, 2006). Variations of the acquaintance immunization scheme have also been used for weighted networks. The acquaintance immunization for weighted networks is outperformed by a strategy in which the immunized neighbors are selected among those with large edge weights (Deijfen, 2011).

A different approach to immunization, the high-risk immunization strategy, applied by Nian and Wang (2010) to the SIRS model, considers a dynamical formulation, in which nodes in contact with one or more infected individuals are immunized with a given probability. Again, by immunizing only a small fraction of the network, a notable reduction of prevalence and increase of the epidemic threshold can be achieved.

Finally, for the SIR model, the mapping to percolation suggests which nodes to target in a vaccination campaign, depending on whether the probability of an outbreak or its size are to be minimized (Kenah and Miller, 2011). A targeted vaccination of nodes in the GSCC implies a reduction of both the probability of a major epidemics and its size.

B. Relevant spreaders and activation mechanisms

Although the problem of immunization is central in the study of epidemics because of its practical implications, the attention of the research community has recently been attracted by the somewhat related theme of discovering which nodes are most influential or effective in the spreading process. For instance, what node should be chosen as initial seed in a SIR epidemic, in order to maximize the total number of nodes eventually reached by the outbreak? This is a very natural question to be posed (Kitsak *et al.*, 2010), in particular, when the propagation process does not involve a disease to be contained but rather a positive meme (such as a crucial piece of information, see Sec. X) whose spreading is instead to be maximized.

The traditional common wisdom, derived from early studies on the immunization problem (Pastor-Satorras and Vespignani, 2002b), was that nodes with the highest degree play the role of superspreaders in networks. This view was challenged by Kitsak *et al.* (2010) who pointed out that the K-core index (see Sec. III.B.5) is a much better predictor of the final outbreak size in the SIR model spreading on several real networks, where (as opposed to uncorrelated networks) the set of nodes with large degrees does not coincide with

high K. The intuitive reason is that the most densely connected core gets easily infected by an outbreak initiated by one of its vertices, finally transmitting the infection to a large portion of the entire network. High degree nodes which are not part of the core may spread the activity to a large number of neighbors but the infection hardly extends farther.

These findings have stimulated a flurry of activity aimed at understanding which of several possible topological centrality measures (degree, betweenness, K-core index, closeness, and many others) are more correlated with spreading influence in various types of networks and contagion dynamics (Bauer and Lizier, 2012; Chen *et al.*, 2012, 2013; Chung *et al.*, 2012; Hou, Yao, and Liao, 2012; P. Li *et al.*, 2012; da Silva, Viana, and da Fontoura Costa, 2012; Hébert-Dufresne *et al.*, 2013; Liu, Ren, and Guo, 2013; Zeng and Zhang, 2013). These studies consider different issues and features of the interplay between the network and the spreading process, and such a large variability does not allow one to reach firm conclusions. Various quantities are used to evaluate the spreading effectiveness: in some cases only top influential spreaders are considered, in others complete rankings of all nodes are compared. Moreover, the consideration of different real networks in different papers does not help in comparing approaches and, in particular, to disentangle the effects of specific topological features such as degree heterogeneity, clustering, or assortativity. Finally not all studies properly take into account the fact that results may be largely different depending on which part of the epidemic phase diagram is considered: the absorbing phase, the transition regime, or the phase where activity is widespread. As a consequence, a clear picture that uniquely determines the best centrality measure that identifies superspreaders for different epidemic models and different networks has yet to be defined.

The K-core decomposition is in many cases a good predictor of spreading efficiency. Nevertheless an interesting finding (Klemm *et al.*, 2012; Hébert-Dufresne *et al.*, 2013) is that the removal of a node with high-K-core index has a limited effect as multiple paths exist among the nodes in the central cores. Thus in general efficient spreaders are not necessarily also good targets for immunization protocols. An extension of the K-core decomposition to weighted networks with application to a SIR epidemics on weighted networks (see Sec. VII.B.3) was also proposed (Garas, Schweitzer, and Havlin, 2012).

Similar to the problem of finding efficient spreaders is the identification of nodes which are infected earlier than the others, thus playing the role of "sensors" for epidemic outbreaks (Christakis and Fowler, 2010; Garcia-Herranz *et al.*, 2014). The strategy of considering friends of randomly chosen nodes allows one to select, without any knowledge of the global network structure, individuals with high degree, high betweenness, small clustering, and high-K-core index, which are actually reached early by epidemic outbreaks. This effect lies at the basis of the acquaintance immunization strategy (Cohen, Havlin, and ben-Avraham, 2003) discussed previously.

Another problem, conceptually close to the search for superspreaders, is the identification of what topological features trigger global epidemics, i.e., what network subsets determine the position of the epidemic threshold (Castellano and Pastor-Satorras, 2012). For SIS, the epidemic threshold scales, within the IBMF approximation, as the inverse of the largest eigenvalue of the adjacency matrix Λ_1 (see Sec. V.A.2). Applying the scaling form of Λ_1 for large uncorrelated scale-free networks (Chung, Lu, and Vu, 2003), the scaling of the threshold with network size is given by Eq. (35). This result can be interpreted as follows (Castellano and Pastor-Satorras, 2012): For $\gamma > 5/2$, the node with the largest degree (hub) together with its direct neighbors forms a self-sustained nucleus of activity above λ_c which propagates to the rest of the system. For $\gamma < 5/2$ instead, the threshold position is dictated by the set of most densely interconnected nodes, as identified by the K-core of the largest index. Topological correlations may alter the picture. For SIR dynamics instead, the largest hub is not able to trigger the transition and the position of the threshold is always dictated by the maximum K-core.

All investigations described so far attempt to relate dynamical properties of the spreading process to purely topological features of the contact pattern. Taking a more general approach, Klemm *et al.* (2012) defined a "dynamical influence" centrality measure, which incorporates not only topological but also dynamical information. The dynamical influence is the leading left eigenvector of a characteristic matrix that encodes the interplay between topology and dynamics. When applied to SIR and SIS epidemic models, the characteristic matrix coincides with the adjacency matrix. The dynamical influence predicts well which nodes are active around the transition, while it is outperformed by other centrality measures far from the threshold (Klemm *et al.*, 2012).

A growing activity has also recently been concerned with the inverse problem of inferring statistically, from the configuration of the epidemics at a given time, which of the nodes was the initial seed originating the outbreak (Comin and da Fontoura Costa, 2011; Pinto, Thiran, and Vetterli, 2012; Brockmann and Helbing, 2013; Altarelli *et al.*, 2014; Lokhov *et al.*, 2014).

Finally, the problem of finding efficient spreaders is not limited to disease epidemic models; it is possibly even more important for complex contagion phenomena (such as rumor spreading or the diffusion of innovations); see Sec. X.

VII. MODELING REALISTIC EPIDEMICS

A. Realistic models

The simple SIS and SIR models considered so far can be generalized to provide a more realistic description of the disease progression by introducing additional compartments (see Sec. II.A) and/or by allowing additional transitions between the different compartments. These variations, that can be studied analytically or most often numerically, may alter the basic phenomenology of the epidemic process. In this section, we survey some of those models and refer the interested reader to the work of Masuda and Konno (2006) for more complicated models that include pathogens' competition and game-theoretical inspired (Webb, 2007) contagion processes.

1. Non-Markovian epidemics on networks

The modeling framework presented in the previous sections is mostly based on the Poisson approximation (Tijms, 2003) for both the transmission and recovery processes. The Poisson approximation assumes that the probabilities per unit time of transmitting the disease through a given edge, or recovering for a given infected node, are constant, and equal to β and μ, respectively. Equivalently, the total time τ_i that a given node i remains infected is a random variable with an exponential distribution $P_i(\tau_i) = \mu e^{-\tau_i \mu}$, and that the time τ_a for an infection to propagate from an infected to a susceptible node along a given edge (the interevent time) is also exponentially distributed $P_a(\tau_a) = \beta e^{-\tau_a \beta}$. A notable variation assumes that all infected nodes remain infective for a fixed time τ. The SIR model can be analyzed exactly in this setting by means of the generating function approach (see Sec. V.B.4).

From a practical point of view, the Poisson assumption leads to an increased mathematical tractability. Indeed, since the rates of transmission and recovery are constant, they do not depend on the previous history of the individual, and thus lead to memoryless, Markovian processes (van Kampen, 1981; Ross, 1996; Tijms, 2003; Van Mieghem, 2014b). While the Poisson approximation may be justified when only the average rates are known (Lambiotte, Tabourier, and Delvenne, 2013), it is at odds with empirical evidence for the time duration of the infective period in most diseases (Blythe and Anderson, 1988), whose distribution usually features a peak centered on the average value but exhibits strongly nonexponential tails. Furthermore, the interest in nonexponential transmission processes has also been fueled by the recent evidence on the patterns of social and communication contacts between individuals, which have been observed to be ruled by heavy-tailed distributions of interevent times (see Sec. VIII).

The framework of non-Poissonian infection and recovery processes can be set up as follows, for either the SIS or SIR model (Boguñá *et al.*, 2014): Infected individuals remain infective for a period of time τ_i, after which they recover, that follows the (nonexponential) $P_i(\tau_i)$ distribution. For simplicity, it is assumed that this distribution is the same for all nodes. Infection events take place along active links, connecting an infected to a susceptible node. Active links transmit the disease at times following the interevent distribution $P_a(\tau_a)$, i.e., a susceptible individual connected to an infected node becomes infected at a time τ_a, measured from the instant the link became active. If a susceptible node is connected to more than one infected node, it becomes infected at the time of the first active link transmitting the disease. The complexity of this non-Markovian process is now evident: the infection of a node depends not only on the number of neighbors, but also on the time at which each connection became active.

Numerical results on non-Poissonian epidemics in networks are relatively scarce. Simple event-driven approaches rely on a time ordered sequence of events (tickets) that represent actions to be taken (recovery or infection) at given fixed times, which are computed from the interevent distributions $P_i(\tau_i)$ and $P_a(\tau_a)$. These approaches are quite demanding, so only small system sizes can be considered. For example, Van Mieghem and van de Bovenkamp (2013) reported results for the SIS model with Poissonian recovery, with rate μ, while infection happens with a nonexponential distribution following the Weibull form $P_a(\tau_a) \sim (x/b)^{\alpha-1} e^{-(x/b)^\alpha}$. In this case, strong variations in the value of the prevalence and of the epidemic threshold are found when varying the parameter α. A promising approach is provided by the general simulation framework proposed by Boguñá *et al.* (2014), based on the extension of the Gillespie algorithm for Poissonian processes (Gillespie, 1977). This algorithm allows the simulation of much larger network sizes.

The consideration of non-Poissonian infection or recovery processes does not lend itself easily to analytical approaches (Lambiotte, Tabourier, and Delvenne, 2013). Some simple forms for the distribution of infectious periods, such as the Erlang distribution, which can be described as the convolution of identical Poisson processes (Cox, 1967), can be tackled analytically by postulating an extended epidemic model with different infective phases and Poissonian transitions among them (Lloyd, 2001a, 2001b). However, general non-Poissonian forms lead to convoluted sets of integrodifferential equations (Keeling and Grenfell, 1997). As a consequence there are not many analytical results for non-Poissonian transitions in complex networks. We mention the results of Min, Goh, and Kim (2013) which consider the SIR process on a network in which infection events follow an interevent distribution $P_a(\tau_a)$. Assuming that infected nodes remain in that state for a fixed amount of time τ_i, it is possible to compute (Min, Goh, and Kim, 2013) the disease transmissibility as

$$T(\tau_i) = 1 - \int_{\tau_i}^{\infty} \psi(\Delta) d\Delta, \tag{67}$$

where $\psi(\Delta) = \int_\Delta^\infty P_a(\tau_a) d\tau_a / \int_0^\infty P_a(\tau_a) d\tau_a$ is the probability distribution of the time between infection (assumed uniform) and the next activation event. Equation (67) assumes that the dynamics of infections follows a stationary renewal process (Cox, 1967; Van Mieghem, 2014b). Applying the generating function approach (see Sec. V.B), the epidemic threshold is obtained, as a function of τ_i, from the implicit equation

$$T(\tau_{ic}) = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{68}$$

For a power-law distribution $P_a(\tau_a) \sim \tau_a^{-\alpha}$, it is found that τ_{ic} diverges as $\alpha \to 2$, implying that only diseases without recovery are able to spread through the network (Min, Goh, and Kim, 2013). An important step forward in the treatment of generic nonexponentially distributed recovery and transmission times in the SIR model is the application of a message-passing method, as reported by Karrer and Newman (2010). This approach leads to an exact description in terms of integrodifferential equations for trees and locally treelike networks, and to exact bounds for non-tree-like networks, in good agreement with simulations.

Finally, Cator, van de Bovenkamp, and Van Mieghem (2013) proposed an extension of the SIS IBMF theory for nonexponential distributions of infection or healing times.

Using renewal theory, their main result is the observation that the functional form of the prevalence in the metastable state is the same as in the Poissonian SIS model, when the spreading rate $\lambda = \beta/\mu$ is replaced by the average number of infection attempts during a recovery time. The theory by Cator, van de Bovenkamp, and Van Mieghem (2013) also allows one to estimate the epidemic threshold in non-Markovian SIS epidemics.

2. The SIRS model

The behavior of the SIRS model on complex networks was analytically considered by Bancal and Pastor-Satorras (2010) at the DBMF level. Within this approximation, the steady-state solution of the SIRS model can be exactly mapped to that of the SIS model, via the identification of the densities of infected individuals

$$\rho_{\mathrm{SIRS}}(\eta,\lambda) = \frac{\eta}{\eta+1}\rho_{\mathrm{SIS}}(\lambda), \tag{69}$$

where η is the immunity decay rate. Therefore, within DBMF, all the critical properties of the SIRS model are the same as the SIS model, the only effect of η being a rescaling of the density of infected individuals.

Numerically, the SIRS model was studied by Abramson and Kuperman (2001) on small-world Watts-Strogatz networks (see Sec. III.D) within a discrete-time deterministic framework, in which infected individuals remain infective for a fixed time τ_I, after which they recover, while recovered individuals remain in this state for a fixed time τ_R. For large values of the Watts-Strogatz model rewiring probability p, a periodic steady state is observed, in which the state of all nodes stays synchronized (Abramson and Kuperman, 2001). The level of synchronization increases with the average degree and also with p, after a threshold p_c depending on $\langle k\rangle$ for fixed network size.

The SIRS model can also be interpreted in terms of a disease that causes death ($I \to R$), leading to an empty node that can be later occupied by the birth of a new, susceptible individual ($R \to S$). Within this interpretation, Liu, Tang, and Yang (2004) considered a generalized SIRS model, allowing additionally for simple recovery ($I \to S$ with rate γ) and death of susceptible individuals due to other causes ($S \to R$ with rate α). Applying a DBMF formalism, they recovered again a threshold inversely proportional to the second moment of the degree distribution, modulated by the diverse parameters in the model, in agreement with the SIS result.

3. The SEIR model

The SEIR model is generally used to model influenzalike illness and other respiratory infections. In the context of networks, this model has been used by Small and Tse (2005) to numerically study the evolution of the severe acute respiratory syndrome (SARS) in different social settings, using both deterministic and stochastic versions of the model, in which different reaction rates were adjusted using empirical spreading data of the disease. The edge-based compartmental modeling approach can be adapted to deal with multiple infectious stages, including SEIR as a particular case (Miller and Volz, 2013).

Exposed individuals can also play a role in more complex epidemiological models. Thus, for example, the SEIRS model can be used to mimic the eventual waning of the immunization of recovered individuals, which implies one additional transition rule, Eq. (4). The properties of the SEIRS model in Watts-Strogatz small-world networks (see Sec. III.D) have been described by Peng and Li (2009). A variation of the SEIRS model without the recovered compartment, or, in other words, in the limit of the reaction rate $\eta \to \infty$ (susceptible-exposed-infected-susceptible), which coincides with a two-stage variation of the classical contact process (Krone, 1999) has been analyzed in heterogeneous networks by Masuda and Konno (2006). Application of DBMF theory recovers the mapping to the simple SIS model obtained in the case of the SIRS epidemics.

B. Realistic static networks

The analytical and numerical results presented so far for the paradigmatic SIS and SIR models have focused mainly on random undirected uncorrelated networks, which are only characterized by their degree distribution, assuming that the rest of the properties are essentially random. However, real networks are far from being completely random. Beyond the degree distribution, a multitude of other topological properties, such as clustering, degree correlations, weight structure, etc. (see Sec. III.A), are needed to characterize them.

1. Degree correlations

Most theoretical results on epidemic spreading in networks, especially at the DBMF level, are obtained imposing a lack of correlations at the degree level, that is, assuming that the probability that a vertex of degree k is connected to a vertex of degree k' is given by $P(k'|k) = k'P(k')/\langle k\rangle$ (Dorogovtsev and Mendes, 2002). However, most natural networks show various levels of correlations, which can have an impact on dynamical processes running on top of them.

From a theoretical point of view, the specific effect of degree correlations, as measured by the different observables detailed in Sec. III.B.3, is difficult to assess. However, some specific results are available. At the level of DBMF theory (see Sec. V.A.1) it was shown that for scale-free networks with $\gamma < 3$ no sort of degree correlations is able to alter the vanishing of the epidemic threshold in the thermodynamic limit (Boguñá, Pastor-Satorras, and Vespignani, 2003a, 2003b). From a numerical point of view, however, the precise determination of the effects of degree correlations on the position of the epidemic threshold and the shape of the prevalence function is problematic. Indeed, it is generally not possible to ascertain if the changes in the epidemic process are due to the presence of correlations or other topological properties generally related to correlations, such as local clustering. Initial simulations on network models (Eguíluz and Klemm, 2002; Warren, Sander, and Sokolov, 2002) claimed that disassortative degree correlations could induce a finite threshold in the SIS model in scale-free networks. However, those claims were based on networks with an underlying finite-dimensional structure (Vázquez *et al.*,

2003), and most probably the finite threshold observed was due to this effect.

For the SIS model, the main IBMF result, Eq. (34), stating that the epidemic threshold is the inverse of the largest eigenvalue of the adjacency matrix Λ_1, remains unaltered. The presence of correlations has only the effect of changing the largest eigenvalue. In this respect, Van Mieghem *et al.* (2010) showed that increasing the degree assortativity, by means of an appropriately defined degree preserving rewiring scheme, increases the largest eigenvalue of the adjacency matrix, thus reducing the effective IBMF epidemic threshold, in a network of fixed size N. On the other hand, the induction of degree disassortativity reduces the largest eigenvalue, with a corresponding increase of the effective IBMF threshold. This observation has been confirmed by Goltsev *et al.* (2012) who estimated, by means of the power iteration method, the largest eigenvalue of the adjacency matrix as

$$\Lambda_1 \simeq \frac{\langle k^2\rangle}{\langle k\rangle} + \frac{\langle k\rangle \sigma^2 r}{\langle k^2\rangle}, \tag{70}$$

where σ is a positive function of the moments of the degree distribution and r is the Pearson correlation coefficient (see Sec. III.B.2). Thus assortativity with $r > 0$ (respectively, disassortativity with $r < 0$) is associated with an increase (respectively, decrease) of the largest eigenvalue. Other properties of the largest eigenvalue in general networks with any kind of correlations, such as the bound

$$\max\left(\sqrt{\langle k^2\rangle}, \sqrt{k_{\max}}\right) \le \Lambda_1 \le k_{\max},$$

are derived by Van Mieghem (2011).

Regarding the SIR model, the mapping to percolation (see Sec. V.B) allows one to obtain more precise information. Assortative correlations can induce a vanishing threshold in networks with finite second moment of the degree distribution (Vázquez and Moreno, 2003). The more general treatment by Goltsev, Dorogovtsev, and Mendes (2008), considering the branching matrix $B_{k,k'} = (k'-1)P(k'|k)$ (Boguñá, Pastor-Satorras, and Vespignani, 2003b), allows one to explicitly check the effects of degree correlations on the epidemic threshold. Indeed disassortative correlations increase the threshold from its uncorrelated value, while assortative correlations decrease it (Goltsev, Dorogovtsev, and Mendes, 2008; Miller, 2009a). These results, valid for the SIR model, can also be extended to the SEIR model (Kenah and Miller, 2011). While no explicit expression for the threshold can be obtained, it is possible to work out upper and lower bounds, in terms of the transmissibility T, that read as

$$\frac{1}{\max_k B(k)} \le T_c \le \frac{\langle k(k-1)\rangle}{\sum_k k(k-1)B(k)P(k)}, \tag{71}$$

where $B(k) = \sum_{k'} B_{k,k'}$ (Goltsev, Dorogovtsev, and Mendes, 2008). With respect to the behavior of the outbreak size close to the epidemic threshold, degree correlations are irrelevant, in the sense that the critical exponents are not changed, when the following conditions are fulfilled (Goltsev, Dorogovtsev, and Mendes, 2008): (i) The largest eigenvalue of the branching matrix is finite if $\langle k^2\rangle$ is finite, and infinite if $\langle k^2\rangle \to \infty$; (ii) the second largest eigenvalue of $B_{k,k'}$ is finite; and (iii) the eigenvector associated with the largest eigenvalue has nonzero components in the limit $k \to \infty$. On the other hand, if any one of these conditions is not fulfilled [large assortativity leads to the failure of condition (ii), while strong disassortativity affects condition (iii)], degree correlations become relevant and they lead to new critical exponents. At the DBMF level the results of Boguñá, Pastor-Satorras, and Vespignani (2003a) for the SIS model extend to the SIR case, implying again the inability of degree correlations to alter the vanishing of the epidemic threshold in the thermodynamic limit for $\gamma < 3$. This result has been confirmed numerically by means of the direct numerical solution of the DBMF equations of the SIR model on scale-free networks with weak assortative correlations (Moreno, Gómez, and Pacheco, 2003). The main effect of these correlations is to induce a smaller overall prevalence and a larger average lifetime of epidemic outbreaks.

2. Effects of clustering

While *a priori* entangled with degree correlations and other topological observables, the effect of clustering on epidemic spreading has been the subject of a large interest, due to the fact that social networks, the basic substrate for human epidemic spreading, are generally highly clustered. Initial work in this area (Keeling, 1999), based on a simple mean-field approximation (and thus valid in principle for homogeneous networks), already pointed out the effects of clustering (measured as the clustering coefficient C, see Sec. III.B.4) on the SIR dynamics. A noticeable departure from the standard mean-field results in the absence of clustering is observed, and, in particular, a decrease of the outbreak size when increasing C. In the case of the Watts-Strogatz model (see Sec. III.D), the paradigm of a network with large clustering, exact analytical results, confirmed by numerical simulations, were obtained by Moore and Newman (2000) for any value of the rewiring probability p. Another analytical approach was proposed by Newman (2003a), who considered a network model based on a one-mode projection of a bipartite network (see Sec. III.C) and applied the usual mapping to percolation. Apart from confirming the observation by Keeling (1999) that epidemic outbreaks are a decreasing function of C, it was observed that, at odds with the behavior of networks with no clustering, for large C the outbreak size saturates to a constant value when increasing the transmissibility even for moderate values of T, suggesting that "in clustered networks epidemics will reach most of the people who are reachable even for transmissibilities that are only slightly above the epidemic threshold" (Newman, 2003a). Along the same line, Miller (2009a), considering a model of random networks with assortative correlations and tunable clustering, was able to show that, for a SIR dynamics with uniform transmissibility T, clustering hinders epidemic spreading by increasing the threshold and reducing prevalence of epidemic outbreaks.

A more general approach, valid for any network, confirms the previous observations (Serrano and Boguñá, 2006). In this approach, the generating function calculation scheme includes

the concept of edge multiplicity m_{ij}, defined as the number of triangles in which the edge connecting nodes i and j participate. In the limit of weak clustering, corresponding to constant $m_{ij} = m_0$, the clustering spectrum (see Sec. III.B.4) follows the scaling $\bar{c}(k) \sim k^{-1}$, which is essentially decoupled from two-vertex degree correlations. The epidemic threshold depends on m_0 and is shifted with respect to the unclustered result; however, for scale-free networks, this shift is not able to restore a finite threshold in the thermodynamic limit. For strong clustering, with a clustering spectrum decaying more slowly than k^{-1}, numerical simulations in a model with tunable clustering coefficient (Serrano and Boguñá, 2005) confirm the inability of clustering to restore a finite threshold in scale-free networks. Other numerical and anaytical works (Miller, 2009a, 2009b) confirmed these results in different clustered network models.

Within the context of IBMF theory for the SIS model, it is possible to find bounds for the largest eigenvalue of the adjacency matrix as a function of the clustering (measured by the number of triangles in the network), indicating that SIS epidemic threshold decreases with increasing clustering coefficient (Van Mieghem, 2011).

3. Weighted networks

If we want to take into account that not all contacts in a social network are equally facilitating contagion (e.g., due to the different relative frequency of physical contacts associated with different edges), we must consider weighted networks, where a weight $\omega_{ij} \geq 0$ is assigned to the edge between connected nodes i and j (see Sec. III.C). The models for epidemic spreading are generalized assuming the rate of disease transmission between two vertices equal to some function of the weight of the link joining them. The simplest possibility occurs when the probability of infection transmission along an edge is directly proportional to the edge weight.

The IBMF theory for the SIS model is readily applied, just replacing in Eq. (33) the adjacency matrix a_{ij} by the matrix $\Omega_{ij} = \omega_{ij} a_{ij}$. The IBMF threshold is the inverse of the largest eigenvalue of Ω (Schumm *et al.*, 2007). Peng, Jin, and Shi (2010) considered a generalized SIS model defined by the matrix β_{ij}, whose terms are the probabilities that node i is infected by node j through an edge joining them. Defining the parametrized adjacency matrix $M_{ij} = \beta_{ij} + (1 - \mu_i)\delta_{ij}$, where μ_i is the recovery probability of node i, Peng, Jin, and Shi (2010) [see also Van Mieghem and Omic (2008)] showed that endemic states occur when the largest eigenvalue (in absolute value) of the parametrized adjacency matrix is larger than 1.

The DBMF approach to the SIS process on weighted networks is simplified by the introduction of additional assumptions, such as a functional dependence of the weights of edges on the degree of the nodes at their end points (Baronchelli and Pastor-Satorras, 2010). Karsai, Juhász, and Iglói (2006) considered the SIS process in a network with local spreading rate, at the DBMF level $\lambda_{kk'} \sim (kk')^{-\sigma}$ with σ in the range [0, 1]. The resulting equations are found to depend on the effective degree exponent $\gamma' = (\gamma - \sigma)/(1 - \sigma)$. For $\gamma' < 3$, a null threshold in the thermodynamic limit is obtained, while for $\gamma' > 3$, the threshold is finite. Karsai, Juhász, and Iglói (2006) additionally discussed a finite-size scaling theory, relating the average prevalence with the network size, which is checked against numerical simulations. The strict correlation between weights and degrees is relaxed in other works, such as Yang and Zhou (2012), where a purely edge-based mean-field approach for weighted homogeneous networks for the SIS model was proposed. By means of this approach, and focusing on bounded and power-law weight distributions, Yang and Zhou (2012) showed that the more homogeneous the weight distribution, the higher is the epidemic prevalence.

Other approaches to the SIS model include a pair-based mean-field approach (Rattana *et al.*, 2013) for networks with random and fixed deterministic weight distributions. The main result is the observation that a weight distribution leads to the concentration of infectiousness on fewer target links (or individuals) which causes an increase in the epidemic threshold in both kinds of networks considered.

Gang *et al.* (2005) reported numerical results for the behavior of the SI model on the growing weighted network model proposed by Barrat, Barthélemy, and Vespignani (2004) with a local spreading rate of the form $\lambda_{ij} \sim (\omega_{ij})^{\alpha}$. The main results obtained concern the slowing down of the disease spread in weighted networks with respect to their unweighted counterparts, which is stronger for larger weight dispersion. Interestingly, they also reported a decay in the velocity of spread, after a sharp peak, taking a slow power-law form, at odds with the exponential form obtained in non-weighted networks (Barthélemy *et al.*, 2005).

In the case of the SIR model Chu *et al.* (2011) presented a DBMF analysis in the case of weights correlated with the degree. The analysis is based on a transmission rate $\lambda_{k'k}$ from vertices of degree k' to vertices of degree k, taking the form $\lambda_{kk'} = \lambda k \omega_{kk'}/s_k$ (where s_k is the strength of a k node) and on an infectivity of nodes $\phi(k)$, denoting the rate at which a node of degree k transmits the disease. Writing down rate equations for the usual relevant DBMF quantities for the SIR model, and assuming $\omega_{kk'} \sim (kk')^{\sigma}$ and $\phi(k) \sim k^{\alpha}$, Chu *et al.* (2011) found the threshold

$$\lambda_c = \frac{\langle k^{\sigma+1} \rangle}{\langle k^{\alpha+\sigma+1} \rangle}. \tag{72}$$

By means of numerical simulations, Chu *et al.* (2011) additionally reported that the size of epidemic outbreaks increases with the exponent α, while it decreases with increasing σ. An analysis of the SIR model in terms of pair approximations for IBMF theory is presented by Rattana *et al.* (2013), reaching analogous results such as those obtained for the SIS model within the same formalism.

It is also noteworthy to mention the numerical work of Eames, Read, and Edmunds (2009) on the SIR model in a realistic social network constructed from actual survey data on social encounters recorded from a peer-group sample of 49 people. The results of Eames, Read, and Edmunds (2009) highlighted the strong correlations between infection risk and node degree and weight, in correspondence with the observations at the DBMF level. Additional simulations considering different immunization strategies (see Sec. VI.A)

indicate that, for this particular realistic network, targeting for total degree or total weight provides approximately the same efficiency levels.

Concerning other models, Britton, Deijfen, and Liljeros (2011) discussed an epidemic model in a weighted network in which the weights attached to nodes of degree k are random variables with probability distributions $q(\omega|k)$, in a construction akin to a weighted configuration model (see Sec. III.D). In this kind of network, Britton, Deijfen, and Liljeros (2011) observed, by means of an analysis based on branching theory, that both the epidemic threshold and the outbreak probability are affected by the correlations between the degree of a node and the weights attached to it. This observation is confirmed by numerical simulations of their weighted network model fitted to empirical data from different network examples, showing that the epidemic threshold is different in the original network with respect to a network with reshuffled weights. On the other hand, Deijfen (2011) analyzed immunization of weighted networks with random and degree dependent weights, observing that targeting the largest weights outperforms other immunization strategies.

In the framework of epidemic models on weighted networks it is possible to also include the CP on networks. In this model each infected node may transmit the disease to at most one neighbor for each time step. This can be interpreted in continuum time as a SIS-like model with a spreading rate $\lambda_{kk'} = 1/k$ for any edge departing from a node of degree k. This modification has the effect of reducing the importance of degree fluctuations in the spreading dynamics: the threshold is finite for any value of the exponent γ (Olinky and Stone, 2004; Castellano and Pastor-Satorras, 2006). The same conclusion can be drawn also for a model where multiple neighbors can be infected simultaneously, but up to a fixed maximum value of neighbors (and not for any k as in SIS) (Joo and Lebowitz, 2004).

4. Directed networks

Directed networks are useful to represent specific types of epidemic transmission in which there is an intrinsic directionality in the propagation. An example is given by diseases communicated by means of blood transfusions or needle sharing. The study of epidemic processes in directed networks is difficult due to the component structure of this kind of networks (see Sec. III.A). Indeed, the position of a node in a specific network component can restrict or enhance its spreading capabilities with respect to other positions. Thus, in order to be able to generate a macroscopic outbreak, a seed of infection should be located on the GIN or GSCC components; seeds on the GOUT or the tendrils will in general produce small outbreaks, irrespective of the spreading rate. In this sense, the distribution of outbreak sizes starting from a randomly chosen vertex is proportional to the distribution of outcomponents.

In the case of the SIR model, the mapping to percolation allows one to apply the generating function formalism developed for percolation in random directed networks (Newman, Strogatz, and Watts, 2001; Schwartz *et al.*, 2002). For purely directed networks (i.e., in which all edges have assigned a directionality), computations depend on the joint probability $P(k^{\text{in}}, k^{\text{out}})$ (see Sec. III.B.2) that a randomly chosen node has in-degree k^{in} and out-degree k^{out}, which in general exhibits correlations between the two values. In the absence of correlations among the degrees of neighbors,[6] under the treelike assumption, the critical transmissibility is

$$T_c = \frac{\langle k^{\text{in}} \rangle}{\langle k^{\text{in}} k^{\text{out}} \rangle}, \tag{73}$$

where averages are taken over the distribution $P(k^{\text{in}}, k^{\text{out}})$ (Newman, Strogatz, and Watts, 2001). The same result can be obtained by means of more intuitive arguments (Schwartz *et al.*, 2002). Equation (73) highlights the important role of correlations between the in-degree and out-degree values in directed networks. Its full discussion is, however, not easy, since one cannot impose arbitrary forms to $P(k^{\text{in}}, k^{\text{out}})$ given the explicit constraint $\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle$. Schwartz *et al.* (2002) discussed the effects of scale-free degree distributions with exponents γ_{in} and γ_{out} for in-degree and out-degree values, respectively, and given correlations $P(k^{\text{in}}, k^{\text{out}})$. With this distribution, epidemics in the GWCC behave as in an undirected network with effective degree distribution $P(k) = \sum_{k^{\text{in}}=0}^{k} P(k^{\text{in}}, k - k^{\text{in}})$, while the β_{SIR} exponent characterizing the size of supercritical outbreaks takes the form of Eq. (19), with an effective $\gamma^* = \gamma_{\text{out}} + (\gamma_{\text{in}} - \gamma_{\text{out}})/(\gamma_{\text{in}} - 1)$ (Schwartz *et al.*, 2002).

More generally, it is possible to consider semidirected networks, in which edges may be directed or undirected (Meyers, Newman, and Pourbohloul, 2006). The network specification is then given in terms of the probability $P(k^{\text{in}}, k^{\text{out}}, k)$ that a vertex has k^{in} incoming edges, k^{out} outgoing edges, and k bidirectional edges. The presence of undirected links implies the existence of short loops of length 2, and thus the violation of the treelike assumption. Considering the possibility of different transmissibilities T_u and T_d for undirected and directed edges, respectively, Meyers, Newman, and Pourbohloul (2006) found expressions for the critical values of one of them, keeping the other fixed. The rather involved expressions simplify when imposing that the in-degree, out-degree, and undirected degree values of each vertex are uncorrelated. In particular, when these quantities obey Poisson distributions, the epidemic threshold is given by (Meyers, Newman, and Pourbohloul, 2006)

$$T_{uc}\langle k \rangle_u + T_{dc}\langle k \rangle_d = 1, \tag{74}$$

where $\langle k \rangle_u$ and $\langle k \rangle_d$ are the undirected and directed average degrees, respectively. The analysis of these results allows the identification of the key epidemiological difference between directed and undirected networks: while in undirected networks the probability of an outbreak and the expected fraction of the population affected (if there is one) are equal, they differ in directed networks: depending on the topology any of the two can be larger (Meyers, Newman, and Pourbohloul, 2006).

[6]Note that these are correlations among two connected vertices, while correlations between k^{in} and k^{out} are for the same node.

The generic case of semidirected networks with arbitrary one-point and two-point correlations is treated in Boguñá and Serrano (2005). The temporal evolution of epidemic outbreaks is considered using the edge-based compartmental modeling in Miller and Volz (2013).

Epidemic processes on purely directed networks can be tackled by an extension of the standard DBMF. The key point is the consideration of new degree classes which are defined in terms of the pair of in-degree and out-degree values $(k^{\text{in}}, k^{\text{out}})$. This implies that the dynamical quantities characterizing the processes also depend on these two values $\rho^{\alpha}_{k^{\text{in}},k^{\text{out}}}$; see Secs. IV.B and V.B.1. Equations for the SIS and SIR models [Eqs. (20) and (47)] translate directly with just one caveat: degree-degree two-vertex correlations (see Sec. III.B.3) in purely directed networks translate into the conditional probability $P^{\text{out}}(k^{\text{in}\prime}, k^{\text{out}\prime}|k^{\text{in}}, k^{\text{out}})$ that an outgoing edge from a vertex $(k^{\text{in}}, k^{\text{out}})$ is connected to a vertex $(k^{\text{in}\prime}, k^{\text{out}\prime})$. Lack of two-point degree-degree correlations implies

$$P^{\text{out}}(k^{\text{in}\prime}, k^{\text{out}\prime}|k^{\text{in}}, k^{\text{out}}) = \frac{k^{\text{in}\prime}P(k^{\text{in}\prime}, k^{\text{out}\prime})}{\langle k^{\text{out}}\rangle}. \tag{75}$$

Boguñá and Serrano (2005) developed this DBMF formalism for the SIR model, finding a threshold that, in the general case, is a function of the largest eigenvalue of the extended connectivity matrix $k^{\text{in}\prime}P(k^{\text{in}\prime}, k^{\text{out}\prime}|k^{\text{in}}, k^{\text{out}})$, and that, without degree-degree correlations, reduces to Eq. (73).

In the case of the SIS model, the IBMF result is the same as in undirected networks, since directionality (i.e., the asymmetry of the adjacency matrix) does not explicitly enter into the theory. See also the generalization of the IBMF theory presented by Peng, Jin, and Shi (2010) (Sec. VII.B.3). The value of the largest eigenvalue has been numerically studied in synthetic semidirected networks with directionality ξ, defined as the fraction of directed edges (Li, Wang, and Van Mieghem, 2013). The main result obtained is the increase of the epidemic threshold lower bound when increasing directionality ξ, implying that directed networks hinder the propagation of epidemic processes. At the DBMF level, an extension analogous to the one considered for the SIR model leads to a threshold with the same functional form, Eq. (73), in degree-degree uncorrelated networks (Tanimoto, 2011).

5. Bipartite networks

Bipartite networks (see Sec. III.C) represent the natural substrate to understand the spreading of sexually transmitted diseases, in which two kinds of individuals (males and females) are present and the disease can be transmitted only between individuals of different kinds.[7] In other contexts, bipartite networks can be used to represent vector-borne diseases, such as malaria, in which the transmission can take place only between the vectors and the hosts (Bisanzio *et al.*, 2010), or the spreading of diseases in hospitals, in which the different kinds of nodes account for (isolated) patients and caregivers (Ancel *et al.*, 2003).

[7]We neglect here homosexual contacts.

Dealing with the SIR dynamics, Newman (2002b) considered a variation of the mapping to percolation, for a model on bipartite networks characterized by the partial degree distributions $P_m(k)$ and $P_f(k)$, finding that the epidemic threshold takes the form of a hyperbola in the space defined by the male and female transmissibilities T_m and T_f,

$$T_m T_f = \frac{\langle k\rangle_m \langle k\rangle_f}{\langle k(k-1)\rangle_m \langle k(k-1)\rangle_f}, \tag{76}$$

where the moments $\langle k\rangle_\alpha$ and $\langle k(k-1)\rangle_\alpha$ are computed for the degree distribution $P_\alpha(k)$.

In the case of the SIS model on bipartite networks, Gomez-Gardenes *et al.* (2008) found analogous results at the DBMF level, with the threshold on the hyperbola defined by the male and female spreading rates λ_m and λ_f of the form

$$\lambda_m \lambda_f = \frac{\langle k\rangle_m \langle k\rangle_f}{\langle k^2\rangle_m \langle k^2\rangle_f}, \tag{77}$$

see Wen and Zhong (2012) for further results with the DBMF formalism. The general behavior of the SIS model on multipartite networks, allowing for more than two different classes of nodes, has been discussed by Santos, Moura, and Xavier (2013).

Expressing in Eq. (76) the transmissibility in terms of the spreading rate $T_i = \lambda_i/(\lambda_i + 1)$ (see Sec. V.B) and comparing with Eq. (77), an interesting observation emerges (Hernández and Risau-Gusman, 2013). In the SIR case, when λ_f diverges the threshold value for λ_m goes to a finite value. Hence the possibility of an endemic outbreak is completely ruled out by reducing the spreading rate of a single type of nodes. In the SIS case instead, the asymptotic value is $\lambda_m = 0$ and as a consequence reducing only one spreading rate may not be sufficient to guarantee no endemic spreading. This last conclusion, however, turns out to be an artifact of the DBMF approach (Hernández and Risau-Gusman, 2013): a finite asymptotic threshold is found also for SIS dynamics in a theoretical approach based on a pair approximation, confirmed by numerical simulations. The previous conclusions hold when the topology-dependent factors appearing on the right-hand sides of Eqs. (76) and (77) are finite. However, it is enough that one of the restricted degree distributions has a diverging second moment to have an epidemics spreading over the whole network, no matter how small the spreading rates λ_i are.

6. Effect of other topological features

Many works have dealt with networks endowed with a modular (community) structure, i.e., subdivided into groups with a relative high density of connections within groups and a smaller density of intergroup links; see Sec. III.B.5. SIS dynamics was studied by Liu and Hu (2005) on a generalization of the classical random graph model with probability p (q) of intra-(inter-)community links. The epidemic threshold is found to decrease with p/q; this effect, however, cannot be attributed to the community structure only, because of the concurrent change of the degree distribution, which gets broader. Other studies have decoupled the two effects, by

comparing spreading dynamics on modular networks and on randomized networks with the same $P(k)$, obtained by suitable reshuffling (Maslov and Sneppen, 2002). They support instead the opposite view that the community structure of a network tends to hinder epidemic spreading. Using IBMF, Bonaccorsi *et al.* (2014) expressed the epidemic threshold explicitly in terms of the sizes and spreading rates in the clusters.

For the SI dynamics, the modular structure makes the growth of the infection slower: prevalence at fixed time is reduced in networks with community structure (Huang and Li, 2007). The interpretation is that the presence of communities tends to confine the outbreak around the initial seed and hinders the transmission to other communities. This effect is further enhanced in weighted social networks (Onnela *et al.*, 2007) by the correlation between topology and weights (Granovetter, 1973): the ties bridging between strongly connected communities are typically weak and this greatly delays the propagation among different communities (Onnela *et al.*, 2007; Karsai *et al.*, 2011). Investigations on the SIRS model with fixed infection and recovery times have focused on the oscillations of the number of infected nodes in the stationary state (Yan *et al.*, 2007; Zhao and Gao, 2007). For both topologies with scale-free and with non-scale-free degree distributions it turns out that the modular structure reduces the synchronization. Also for SIR dynamics modularity is found to make spreading more difficult: the final value of ρ^R is smaller for stronger community structure (Wu and Liu, 2008). More convincingly, Salathé and Jones (2010) showed, for both empirical and synthetic networks, that community structure has a major hindering effect on spreading: the final value of ρ^R and the height of the peak of ρ^I decrease with the modularity. Moreover, they showed that in networks with strong community structure targeting vaccination interventions at individuals bridging communities is more effective than simply targeting highly connected individuals.

It is also worth mentioning the observation that SIS-like processes on complex networks may give rise to the nontrivial scenario of Griffiths phases (Vojta, 2006), regions of the phase space where the only stationary state is the absorbing one, which is however reached via anomalously long nonuniversal relaxation (Muñoz *et al.*, 2010). This behavior arises because of rare-region effects, which can be due either to quenched local fluctuations in the spreading rates or to subtle purely topological heterogeneities (Juhász *et al.*, 2012; Ódor and Pastor-Satorras, 2012). Such rare-region effects have been discussed in the case of the SIS model on loopless (tree) weighted networks (Buono *et al.*, 2013; Ódor, 2013a, 2013b), where they have been related to the localization properties of the largest eigenvalue of the adjacency matrix (Ódor, 2013b).

7. Epidemics in adaptive networks

Previous sections focused on the evolution of epidemics on static networks or on annealed topologies where connections are rewired on a time scale much smaller than the characteristic time scale of the infection process. For real human disease epidemics, however, the assumption that the structure of contacts does not depend on the progression of the contagion is often unrealistic: In the presence of infectious spreading, human behavior tends to change spontaneously, influencing the spreading process itself in a nontrivial feedback loop. The modifications induced by this coupling may be distinguished depending on several features (Funk, Salathé, and Jansen, 2010): the source of information about the contagion, the type of information considered, and the type of behavioral change induced. The source of information about the spreading process may be local (individuals decide depending on the state of their direct contacts) or global (information on the state of the whole system is publicly available). Different types of information may influence the behavioral choice: in prevalence-based models decisions are made based on the observation of the epidemic state of others; in belief-based models what matters is the awareness or the risk perception which may be (at least partially) independent from the actual disease dynamics and often behaves in turn as a spreading process (Bagnoli, Liò, and Sguanci, 2007; Salathé and Bonhoeffer, 2008; Funk *et al.*, 2009; Perra *et al.*, 2011; Bauch and Galvani, 2013; Granell, Gómez, and Arenas, 2013). Finally, the behavioral change can be of different types: affecting the state of the individual (for example, via voluntary vaccination) or the structure of contacts (eliminating existing connections or creating new ones). Many models incorporating these features have been investigated in mathematical epidemiology, generally assuming well-mixed populations (Funk, Salathé, and Jansen, 2010). Here we focus on epidemic spreading on adaptive (or coevolving) contact networks, where the topology of the interaction pattern changes in response to the contagion. The coevolution between structure and dynamics is a common theme in many contexts, from game theory to opinion dynamics (Gross and Blasius, 2008; Nardini, Kozma, and Barrat, 2008).

The first investigation of an adaptive topology for SIS dynamics (Gross, D'Lima, and Blasius, 2006) includes the possibility for individuals to protect themselves by avoiding contacts with infected people. Infected individuals are allowed at each time step to infect each of their susceptible contacts with probability p or recover with probability r (usual SIS dynamics); in addition, susceptibles can decide (with probability w) to sever a link with an infected and reconnect to a randomly chosen susceptible. The possibility of rewiring links drastically changes the phase diagram of the model. The threshold p_c, below which the system always converges to the absorbing healthy state, is much larger than in the case of no coevolution ($w = 0$): rewiring hinders the disease propagation. More interestingly, above this threshold a bistability region appears (see Fig. 7) with associated discontinuous transitions and hysteresis. In this region both the healthy and endemic states are stable and the fate of the system depends on the initial condition. If p is further increased above a second threshold, bistability ends and the endemic state is the only attractor of the dynamics. The coevolution also has strong effects on the topology of the contact network, leading to the formation of two loosely connected clusters of infecteds and susceptibles, with a general broadening of the degree distribution and buildup of assortative correlations. The rich phase diagram is recovered by a simple homogeneous mean-field approach which complements the equation for the prevalence with two additional equations for the density of links of *I*-*I* and *S*-*I* types. A bifurcation analysis also predicts

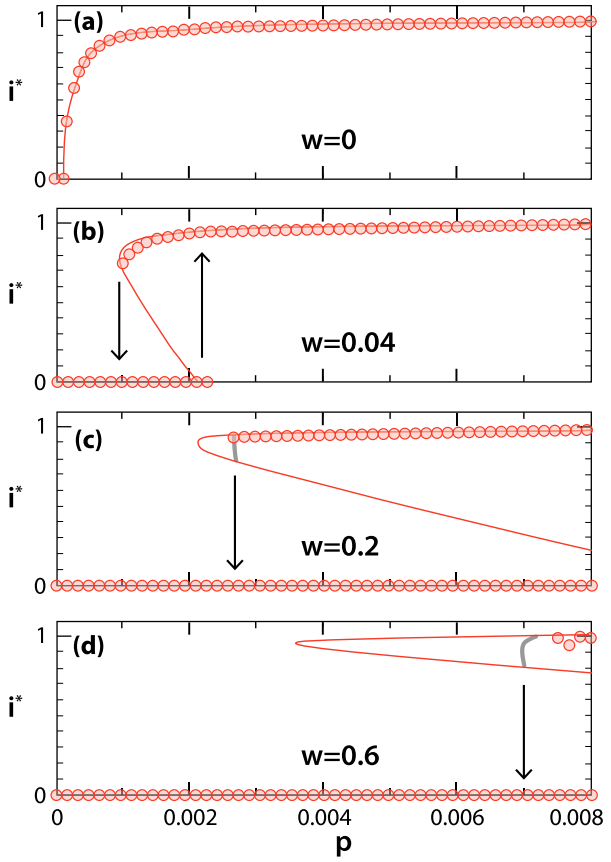

FIG. 7 (color online). Density of the infected nodes i^* as a function of the infection probability p for different values of the rewiring rate w. In each diagram thin lines are computed using a homogeneous mean-field approach while circles are the results of numerical simulations. Without rewiring only a single continuous transition occurs for $p_c \approx 0.0001$ (a). By contrast, rewiring causes a number of discontinuous transitions, bistability, and hysteresis loops (indicated by arrows) in (b)–(d). Adapted from Gross, D'Lima, and Blasius, 2006.

the existence of a very narrow region with oscillatory dynamics. A more detailed approach to the same dynamics (Marceau *et al.*, 2010) explicitly takes into account the degree of nodes, writing equations for the evolution of the probabilities S_{kl} (I_{kl}) that nodes in state S (I) have degree k and l infected neighbors. The numerical integration of the equations is in excellent agreement with numerical simulations with respect to both the transient evolution and the stationary state. Different initial topologies (degree-regular, Poisson, power-law distributed) with the same average connectivity may lead to radically different stationary states: either fully widespread contagion or rapid disease extinction.

The qualitative picture emerging from the model of Gross, D'Lima, and Blasius (2006) is found also for the adaptive SIRS model (Shaw, 2008) and for the SIS dynamics where a susceptible individual rewires to any randomly chosen other vertex (not necessarily susceptible) (Zanette and Risau-Gusmán, 2008). The possibility that also infected individuals decide to rewire their connections has been discussed by Risau-Gusman and Zanette (2009). In the SIS model, the interplay of the adaptive topology and vaccination has also been investigated (Shaw and Schwartz, 2010). It turns out that the vaccination frequency needed to significantly lower the disease prevalence is much smaller in adaptive networks than in static ones.

The effect of the same type of adaptive rewiring introduced for SIS has also been studied for SIR dynamics (Lagorio *et al.*, 2011). In this case the effects of the coevolution are less strong, as the time needed to reach the stationary (absorbing) state is short (logarithmic in the system size N) and the global topology is only weakly perturbed in this short interval. The phase diagram remains qualitatively the same of the non-adaptive case with a single epidemic transition separating a healthy state from an endemic one. The mapping to percolation (see Sec. V.B) is also useful here. Coevolution leads to an effective transmissibility T which decreases with the rewiring probability w. One can then identify a critical value w_c above which the adaptive behavior is sufficient to completely suppress the epidemics.

The assumptions that disconnected links are immediately rewired and that the target vertices of the reconnection step are randomly selected in the whole network are highly implausible in real-world situations. Attempts to go beyond these limitations include the consideration of different rates for breaking and establishing links (Van Segbroeck, Santos, and Pacheco, 2010; Guo *et al.*, 2013) and "intermittent" social distancing strategies, such that a link is cut and recreated (between the same vertices) after a fixed time interval (Valdez, Macri, and Braunstein, 2012a) or with a certain rate after both end points have healed (Tunc, Shkarayev, and Shaw, 2013). The latter strategies are intended to mimic what happens with friends or working partners, with which connections are reestablished after the disease. The overarching structure of the network remains static and there is no real coevolution (no new links are formed). As a consequence the phase diagram of epidemic models remains the same found on static networks, with only an increase in the epidemic threshold due to social distancing.

C. Competing pathogens

Another generalization of the basic modeling scheme considers the evolution of multiple epidemic processes in competition in the same network, a scenario with clear relevance for realistic situations. The crucial concept here is cross immunity, i.e., the possibility that being infected by one pathogen confers partial or total immunity against the others.

Newman (2005) considered two SIR epidemic processes occurring one after the other in the same static network, in conditions of total cross immunity: The second pathogen can affect only survivors of the first, i.e., in the "residual" network obtained once the nodes recovered when the first epidemics ends are removed. The mapping of SIR static properties to bond percolation allows one to understand this case. If the first pathogen is characterized by a transmissibility above a certain value (coexistence threshold), the residual network has no giant component and the second pathogen cannot spread globally, even if it has a large transmissibility. Global spreading of both pathogens can occur only for values of the transmissibility of the first infection in an interval between

the epidemic and the coexistence thresholds. A generalization to the case of partial cross immunity has been discussed by Funk and Jansen (2010). The case of competing SIR infections spreading concurrently has been investigated by Karrer and Newman (2011), again in the case of complete cross immunity: Infection by one pathogen confers immunity for both. Nontrivial effects occur when both transmissibilities are above the threshold for single spreading (otherwise one of the pathogens does not spread globally and there is no real interference). If one of the pathogens has a transmissibility significantly larger than the other, it spreads fast and the second spreads afterward in the residual network, much as in the case of subsequent infections. If the growth rates are similar, the final outcome shows strong dependence on stochastic fluctuations in the early stages of growth, with very strong finite-size effects. An alternative approach, based on the edge-based compartmental modeling, allows one to theoretically investigate the dynamics of two competing infectious diseases (Miller, 2013). Poletto *et al.* (2013) considered cross-immune pathogens in competition within a metapopulation framework (see Sec. IX). The dominance of the strains also depends in this case on the mobility of hosts across different subpopulations.

Mutual cross immunity for two competing SIS dynamics has been considered by Trpevski, Tang, and Kocarev (2010) [see also Ahn *et al.* (2006)], while the domination time of two competing SIS viruses was analyzed by van de Bovenkamp, Kuipers, and Van Mieghem (2014). Depending on the network topology, for some values of the parameters it is possible to find a steady state where the two processes coexist, each having a finite prevalence.

Another nontrivial and relevant example of interacting epidemics is the case of coinfection processes, where the opposite of cross immunity holds: The second pathogen can spread only to individuals that have been already infected by the first. Newman and Ferrario (2013) reported the first theoretical and numerical investigation of this type of dynamics on complex networks.

VIII. EPIDEMIC PROCESSES IN TEMPORAL NETWORKS

The majority of the results presented so far considered the spreading of the epidemic process in the limit of extreme time scale separation between the network and the contagion process dynamics (see, however, Sec. VII.B.7 for a discussion on adaptive networks, whose topology changes in reaction to a disease). In static networks, the epidemic spreads on a network that is virtually frozen on the time scale of the contagion process. In the opposite limit, the DBMF theory considers an effective mean-field network where nodes are effectively rewired on a time scale much faster than the contagion process. However, in the case of many real-world networks those assumptions are rather simplistic approximations of the real interplay between time scales. For instance, in social networks, no individual is in contact with all of his or her friends simultaneously all the time. On the contrary, contacts are changing in time, often on a time scale that is comparable with the one of the spreading process. Real contact networks are thus essentially dynamic, with connections appearing, disappearing, and being rewired with different characteristic time scales, and are better represented in terms of a temporal or time-varying network (Holme and Saramäki, 2012, 2013), see Fig. 8.

Temporal networks are defined in terms of a *contact sequence*, representing the set of edges present at a given time t. By aggregating the instantaneous contact sequence at all times $t < T$, a static network projection can be constructed; see Fig. 8. In this aggregated network, the edge between nodes i and j is present if it ever appeared at any time $t < T$. A more informative static representation is a weighted network, in which the weight associated with each edge is proportional to the total number of contacts (or the total amount of time the contact was active) between each pair of individuals. These static network projections, however, do not account for the nontrivial dynamics of the temporal network and are thus often inappropriate when considering dynamical processes unfolding on time-varying connectivity patterns.

Recent technological advances allow one to gather large amounts of data on social temporal networks, such as mobile phone communications (Onnela *et al.*, 2007) and face-to-face interactions (Cattuto *et al.*, 2010). From the analysis of these data sets, social interactions are characterized by temporally heterogeneous contact patterns.

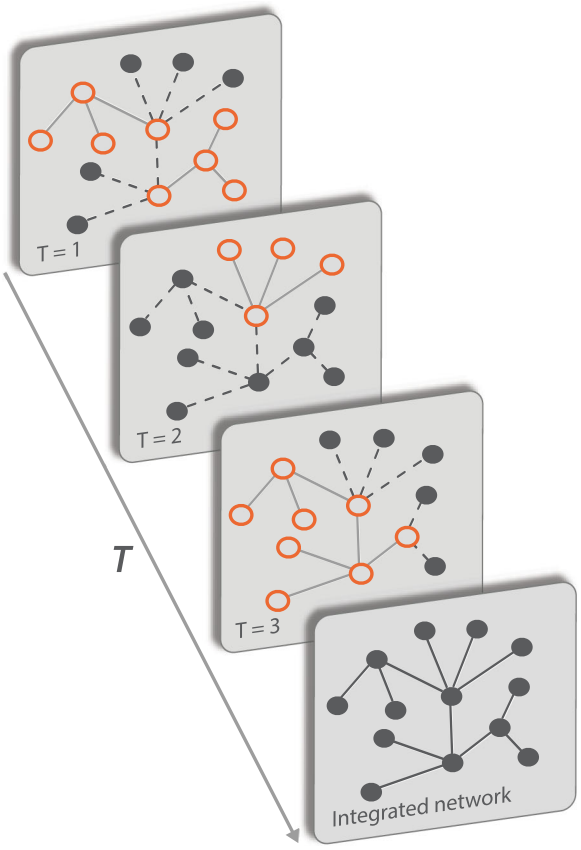

FIG. 8 (color online). A temporal (or time-varying) network can be represented as a set of nodes that, at every instant of time, are connected by a different set of edges. An integrated network over a time window T is constructed by considering that nodes i and j are connected by an edge if they were ever connected at any time $t \leq T$. Adapted from Perra, Gonçalves *et al.*, 2012.

Indeed it is more the norm than the exception to find that the temporal behavior of social interactions is characterized by heavy-tail and skewed statistical distributions. For instance, the probability distributions of the length of contacts between pairs of individuals, of times between consecutive interactions involving the same individual, etc., all follow a heavy-tailed form (see Fig. 9) (Holme, 2005; Hui *et al.*, 2005; Onnela *et al.*, 2007; Cattuto *et al.*, 2010; Tang *et al.*, 2010; Doerr, Blenn, and Van Mieghem, 2013). These properties contrast with the Poissonian behavior expected in purely random interactions, thus catalyzing the recent interest in the study of the *burstiness* of human behavior (Oliveira and Barabasi, 2005).

The time-varying connectivity pattern of networks affects epidemic processes in a number of different ways. The presence of a temporal ordering in the connections of the network limits the possible paths of propagation of the epidemic process. In particular, not all the edges of the eventually aggregated network projection are available for the propagation of a disease. Starting on a given node, only the nodes that belong to its *set of influence* (Holme, 2005), defined as the nodes that can be reached through paths that respect time ordering, may propagate the disease. Furthermore, the Poissonian approximation for the transmission rate of infectious individuals is not correct because the time between consecutive nodes' contacts is generally

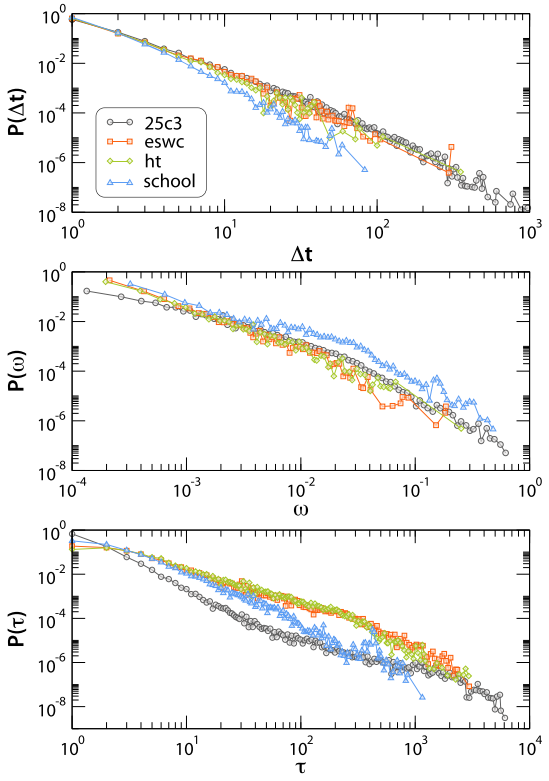

FIG. 9 (color online). Statistical properties of four temporal face-to-face contact networks (Cattuto *et al.*, 2010). The probability distributions of the length of conversations Δt, total time spent in conversation between pairs of individuals ω, and the gap τ between conversation with different individuals all show a long-tailed form, compatible with a power law. Adapted from Starnini *et al.*, 2012.

power-law distributed. However, this non-Poissonian behavior is different from the one presented in Sec. VII.A.1, where we considered fixed networks in which a disease takes, to propagate from an infected individual to a susceptible one along a fixed link, a time τ_a that is not exponentially distributed. Here we have the situation in which the very link that can propagate the disease appears at instants of time that are separated by an interevent time τ_l that can be distributed nonexponentially. Finally, the relation between the intrinsic time scales of the temporal network and those of the dynamics plays a substantial role. Thus, for slow dynamics with a very large relative time scale, it can be a good approximation to consider as a substrate the weighted integrated network. If the dynamics is fast, with a small relative time scale, comparable to that of the temporal network, then the substrate must be the actual contact sequence defining the temporal network.

Among the effects that a non-Poissonian temporal network induces on epidemic spreading, one of the most remarkable is a substantial slowing down of the spread velocity. This observation was first made by using an SI model (Vazquez *et al.*, 2007) [see also Min, Goh, and Vazquez (2011)] in the context of the spreading of email worms among email users. Empirical data show that the distribution of times between consecutive email activities is heavy tailed and well approximated by the form $P(\tau') \sim \tau'^{-1-\beta}$. The generation time τ, defined as the time between the infection of the primary individual and the infection of a secondary individual, is given by the residual waiting time distribution, assuming a stationary process (Cox, 1967) $g(\tau) = \int_\tau^\infty P(\tau')d\tau'/\langle\tau\rangle \sim \tau^{-\beta}$, where it is assumed that the time at which emails are received is uniformly random. The average number of new infections at time t, $n(t)$ is estimated as $n(t) = \sum_{d=1}^{D} Z_g \hat{g}_d(t)$, where Z_d is the average number of users at a distance d (at d email steps) from the first infected user, D is the maximum possible value of d, and $\hat{g}_d(t)$ is the convolution of order d of $g(\tau)$. Assuming that the integrated network of email contacts is sparse, Min, Goh, and Vazquez (2011) found that $n(t) \sim t^{-\beta}$, independently of the integrated network structure. This result implies that the disease spreads much more slowly than in a regular static network, where an exponential increase of infected individuals is observed. The slowing down in temporal networks has been empirically measured in different systems (Vazquez *et al.*, 2007; Karsai *et al.*, 2011; Stehle *et al.*, 2011; Kivelä *et al.*, 2012) and also reported in other dynamical processes, such as diffusion (Hoffmann, Porter, and Lambiotte, 2012; Perra, Baronchelli *et al.*, 2012; Starnini *et al.*, 2012) or synchronization (Fujiwara, Kurths, and Díaz-Guilera, 2011). The situation is however not completely clear, since other works suggest instead a dynamic acceleration (Jo *et al.*, 2014). These temporal effects are, moreover, entangled with topological ones, as shown by Rocha, Liljeros, and Holme (2011) analyzing the SI and SIR models in empirical spatiotemporal networks. Temporal correlations accelerate epidemic outbreaks, especially in the initial phase of the epidemics, while the network heterogeneity tends to slow them down.

The time-varying structure of temporal networks is also able to alter the value of the epidemic threshold, as analytically shown for the SIS and SIR processes in activity-driven

network models (Perra, Gonçalves *et al.*, 2012). The activity-driven network class of models (Perra, Gonçalves *et al.*, 2012; Starnini and Pastor-Satorras, 2013) is based on the concept of activity potential, defined as the probability per unit time that an individual engages in a social activity. Empirical evidence shows that the activity potential varies considerably from individual to individual and the dynamics of the networks is encoded in the function $F(a)$ that characterizes the probability for a node to have an activity potential a. The activity-driven network model considers N nodes whose activity a_i is assigned randomly according to the distribution $F(a)$. During each time step the node i is considered active with probability a_i. Active nodes generate m links (engage in m social interactions) that are connected to m individuals chosen uniformly at random. Finally, time is updated $t \to t+1$. The model output is a sequence of graphs, depending on the distribution $F(a)$, which is updated at every time step t. An integrated network at time T can be constructed by considering the union of the sequence of graphs; see Fig. 8. This integrated network has a degree distribution which depends on the activity distribution as $P_T(k) \simeq (1/T)F(k/T - \langle a \rangle)$ (Starnini and Pastor-Satorras, 2013), where $\langle a \rangle$ is the average activity and for simplicity we take $m = 1$. The empirically observed power-law activity distributions $F(a)$ can thus explain the long tails in the degree distribution of social networks (Perra, Gonçalves *et al.*, 2012). Perra, Gonçalves *et al.* (2012) considered the behavior of the SIS model in activity-driven networks, writing dynamical mean-field equations for the infected individuals in the class of activity rate a, at time t, namely, I_a^t. The discrete-time dynamical evolution considers concurrently the dynamics of the network and the epidemic model, yielding

$$I_a^{t+1} = \lambda m(N_a - I_a^t)a \int da' \frac{I_{a'}^t}{N} + \lambda m(N_a - I_a^t) \int da' \frac{I_{a'}^t a'}{N}, \tag{78}$$

where $N_a = F(a)N$ is the total number of individuals with activity a and where the recovery probability $\mu = 1$. In Eq. (78), the first term on the right-hand side takes into account the probability that a susceptible of class a is active and acquires the infection getting a connection from any other infected individual (summing over all different classes), while the last term takes into account the probability that a susceptible, independently of his activity, gets a connection from any infected active individual. A linear stability analysis of Eq. (78) leads to an epidemic threshold

$$\lambda_c = \frac{1}{m(\langle a \rangle + \sqrt{\langle a^2 \rangle})}. \tag{79}$$

The same epidemic threshold is obtained for the SIR model, applying mean-field approximations (Liu *et al.*, 2014) and a mapping to percolation (Starnini and Pastor Satorras, 2014). This result highlights the crucial fact that scale-free integrated networks can lead to a vanishing threshold for epidemics with a very large time scale, while epidemics with a short time scale, comparable to the one of the contact sequence, can be associated with a finite, nonvanishing threshold; see Fig. 10.

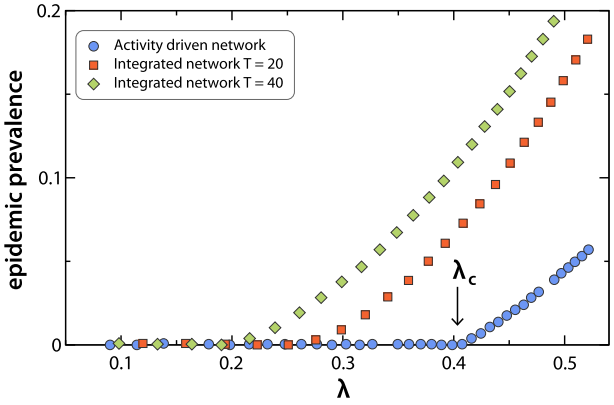

FIG. 10 (color online). Prevalence of the SIS model on the temporal network defined by the activity-driven model, as a function of the basic transmission probability λ. The threshold observed for the dynamics on the temporal network coincides with the theoretical prediction Eq. (79). Simulations on integrated networks show instead a threshold that becomes smaller when increasing the integration time T. Adapted from Perra, Gonçalves *et al.*, 2012.

This observation has been confirmed in studies of other temporal network models (Rocha and Blondel, 2013).

Finally, a recent avenue of research in this area has been the identification of effective immunization protocols for temporal networks (Lee *et al.*, 2012). The idea here is to define a training window ΔT, such that information is gathered from the contact sequence at times $t < \Delta T$. A set of individuals to be immunized is chosen and effectively vaccinated at time ΔT. The effects of the immunization are then observed for $t > \Delta T$. Lee *et al.* (2012) explored two local strategies, inspired by the acquittance immunization protocol for static networks (Cohen, Havlin, and ben-Avraham, 2003): In the "recent" strategy, a randomly chosen individual is asked at time ΔT for its last contact; this last contact is immunized. In the "weight" strategy, a randomly chosen individual at time ΔT is asked for its most frequently contacted peer, up to time ΔT; this most frequent contact is immunized. By means of numerical simulations Lee *et al.* (2012) observed that both protocols offer, for a limited amount of local information, a reasonable level of protection against the disease propagation. An interesting issue is the question about the amount of information (the length ΔT of the training window) sufficient to achieve an optimal level of immunization for a fixed fraction of immunized individuals. Starnini *et al.* (2013) found a saturation effect of the level of immunization for training windows of about 20%–40% of the total length of the contact sequence, for several immunization protocols, indicating that a limited amount of information is actually enough to optimally immunize a temporal network. In the case of the activity-driven networks, analytical expressions for several immunization strategies can be obtained (Liu *et al.*, 2014).

IX. REACTION-DIFFUSION PROCESSES AND METAPOPULATION MODELS

So far we reviewed results concerning spreading and contagion processes in which each node of the network

corresponds to a single individual of the population. A different framework emerges if we consider nodes as entities where multiple individuals or particles can be located and eventually wander by moving along the links connecting the nodes. Examples of such systems are provided by mechanistic epidemic models where particles represent people moving between different locations or by the routing of information packets in technological networks (Sattenspiel and Dietz, 1995; Keeling and Rohani, 2002; Gallos and Argyrakis, 2004; Watts *et al.*, 2005). More in general models of social behavior and human mobility are often framed as reaction-diffusion processes where each node i is allowed to host any nonnegative integer number of particles $\mathcal{N}(i)$, so that the total particle population of the system is $\mathcal{N} = \sum_i \mathcal{N}(i)$. This particle-network framework considers that each particle diffuses along the edges connecting nodes with a diffusion coefficient that depends on the node degree and/or other node attributes. Within each node particles may react according to different schemes characterizing the interaction dynamics of the system. A simple sketch of the particle-network framework is represented in Fig. 11.

In order to have an analytic description of reaction-diffusion systems in networks one has to allow the possibility of heterogeneous connectivity patterns among nodes. A first analytical approach to these systems considers the extension of the degree-based mean-field approach to reaction-diffusion systems in networks with arbitrary degree distribution. For simplicity, we first consider the DBMF approach to the case of a simple system in which noninteracting particles (individuals) diffuse on a network with arbitrary topology. A convenient representation of the system is therefore provided by quantities defined in terms of the degree k:

$$\mathcal{N}_k = \frac{1}{N_k} \sum_{i \in \mathcal{V}(k)} \mathcal{N}(i), \tag{80}$$

where $N_k = NP(k)$ is the number of nodes with degree k and the sum runs over the set of nodes $\mathcal{V}(k)$ having degree equal to k. The degree block variable $\mathcal{N}_k$ represents the average number of particles in nodes with degree k. The use of the DBMF approach amounts to the assumption that nodes with degree k, and thus the particles in those nodes, are statistically

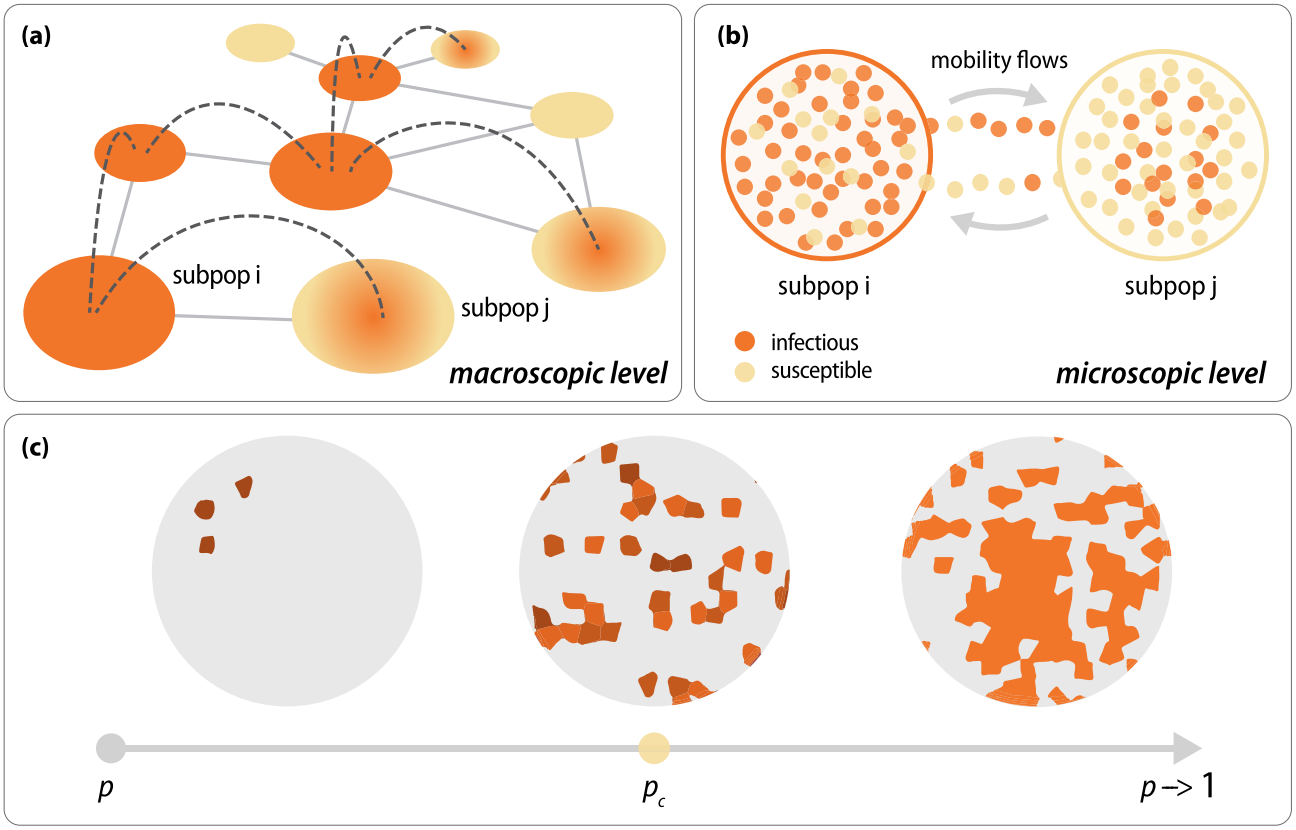

FIG. 11 (color online). (a) Schematic illustration of the simplified modeling framework based on the particle-network scheme. At the macroscopic level the system is composed of a heterogeneous network of subpopulations. The contagion process in one subpopulation can spread to other subpopulations because of particles diffusing across subpopulations. (b) At the microscopic level, each subpopulation contains a population of individuals. The dynamical process, for instance, a contagion phenomenon, is described by a simple compartmentalization (compartments are indicated by different colored dots in the picture). Within each subpopulation, individuals can mix homogeneously or according to a subnetwork and can diffuse with probability p from one subpopulation to another following the edges of the network. (c) A critical value p_c of the individuals or particles diffusion identifies a phase transition between a regime in which the contagion affects a large fraction of the system and one in which only a small fraction is affected (see the discussion in the text).

equivalent. In this approximation the dynamics of particles randomly diffusing on the network is given by a mean-field dynamical equation expressing the variation in time of the particle subpopulation $\mathcal{N}_k(t)$ in each degree block k. This can be easily written as

$$\frac{d\mathcal{N}_k}{dt} = -d_k\mathcal{N}_k(t) + k\sum_{k'}P(k'|k)d_{k'k}\mathcal{N}_{k'}(t). \tag{81}$$

The first term of the equation considers that only a fraction d_k of particles moves out of the node per unit time. The second term instead accounts for the particles diffusing from the neighbors into the node of degree k. This term is proportional to the number of links k times the average number of particles coming from each neighbor. This is equal to the average over all possible degrees k' of the fraction of particles moving on that edge, $d_{k'k}\mathcal{N}_{k'}(t)$, according to the conditional probability $P(k'|k)$ that an edge belonging to a node of degree k is pointing to a node of degree k'. Here the term $d_{k'k}$ is the diffusion rate along the edges connecting nodes of degree k and k'. The rate at which individuals leave a subpopulation with degree k is then given by $d_k = k\sum_{k'}P(k'|k)d_{kk'}$. In the simplest case of homogeneous diffusion each particle diffuses with rate r from the node in which it is and thus the diffusion per link $d_{k'k} = r/k'$. On uncorrelated networks $P(k'|k) = k'P(k')/\langle k\rangle$ and hence one easily gets in the stationary state $d\mathcal{N}_k/dt = 0$ the solution (Noh and Rieger, 2004; Colizza, Pastor-Satorras, and Vespignani, 2007)

$$\mathcal{N}_k = \frac{k}{\langle k\rangle}\frac{\mathcal{N}}{N}. \tag{82}$$

Equation (82) explicitly brings the diffusion of particles in the description of the system and points out the importance of network topology in reaction-diffusion processes. This expression indicates that the larger the degree of a node, the larger the probability to be visited by the diffusing particles.

A. SIS model in metapopulation networks

The above approach can be generalized to reacting particles with different states by adding a reaction term to the above equations (Colizza, Pastor-Satorras, and Vespignani, 2007). We now describe a generalization to this setting of the standard SIS model in discrete time, with probability per unit time β of infection and probability μ of recovery. We consider $\mathcal{N}$ individuals diffusing in a heterogeneous network with N nodes and degree distribution $P(k)$. Each node i of the network has a number $I(i)$ of infectious and $S(i)$ of susceptible individuals. The occupation numbers $I(i)$ and $S(i)$ can have any integer value, including $I(i) = S(i) = 0$, that is, void nodes with no individuals. This modeling scheme describes spatially structured interacting subpopulations, such as city locations, urban areas, or defined geographical regions (Grenfell and Harwood, 1997; Hanski and Gaggiotti, 2004) and is usually referred to as the *metapopulation approach*. Each node of the network represents a subpopulation and the compartment dynamics accounts for the possibility that individuals in the same location may get into contact 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. We have thus to associate with each individual's class a diffusion probability p_I and p_S that indicates the probability for any individual to leave its node and move to a neighboring node of the network. In general the diffusion probabilities are heterogeneous and can be node dependent; however, for simplicity we assume that individuals diffuse with probability $p_I = p_S = 1$ along any of the links departing from the node in which they are. This implies that at each time step an individual sitting on a node with degree k will diffuse into one of its nearest neighbors with probability $1/k$. In order to write the dynamical equations of the system we define the following quantities:

$$I_k = \frac{1}{N_k}\sum_{i\in\mathcal{V}(k)}I(i), \qquad S_k = \frac{1}{N_k}\sum_{i\in\mathcal{V}(k)}S(i), \tag{83}$$

where the sums $\sum_{i\in\mathcal{V}(k)}$ are performed over nodes of degree k. These two quantities express the average number of susceptible and infectious individuals in nodes with degree k. Clearly, $\mathcal{N}_k = I_k + S_k$ is the average number of individuals in nodes with degree k. These quantities allow one to write the discrete-time equation describing the time evolution of $I_k(t)$ for each class of degree k as

$$I_k(t+1) = k\sum_{k'}P(k|k')\frac{1}{k'}[(1-\mu)I_{k'}(t) + \beta\Gamma_{k'}(t)], \tag{84}$$

where $\Gamma_{k'}(t)$ is an interaction kernel, a function of $I_{k'}$ and $S_{k'}$. Equation (84) is obtained by considering that at each time step the particles present on a node of degree k, first react, and then diffuse away from the node with probability 1. The value of $I_k(t+1)$ is obtained by summing the contribution of all particles diffusing to nodes of degree k from their neighbors of any degree k', including the new particles generated by the reaction term $\Gamma_{k'}$. In the case of uncorrelated networks, Eq. (84) reduces to

$$I_k(t+1) = \frac{k}{\langle k\rangle}[(1-\mu)\bar{I}(t) + \beta\Gamma], \tag{85}$$

where $\bar{I}(t) = \sum_k P(k)I_k$ is the average number of infected individuals per node in the network and $\Gamma = \sum_k P(k)\Gamma_k$. Analogously the equation describing the dynamics of susceptible individuals is

$$S_k(t+1) = \frac{k}{\langle k\rangle}[\bar{S}(t) + \mu\bar{I}(t) - \beta\Gamma], \tag{86}$$

where $\bar{S}(t) = \sum_k P(k)S_k$.

In order to explicitly solve these equations we have to specify the type of interaction among individuals. In the usual case of a mass-action law for the force of infection, we have $\Gamma_k = I_kS_k/\mathcal{N}_k$. This implies that each particle has a finite number of contacts with other individuals. Considering the

stationary state $t \to \infty$, and by using some simple algebra, we can find that an endemic state $\bar{I} > 0$ occurs only if $\beta/\mu > 1$, thus recovering the classic epidemic threshold in homogeneous systems (Colizza, Pastor-Satorras, and Vespignani, 2007).

A different result is obtained if we consider the case in which each susceptible individual may react with all the infectious individuals in the same node. In this case $\Gamma_k = I_k S_k$, i.e., all individuals are in contact with the same probability (absorbed in the factor β), independently of the total population present in each node. This law, referred to as the pseudomass-action law, is sometimes used to model animal diseases as well as mobile phone malwares. In this case, an active stationary solution $\bar{I} > 0$ occurs if (Colizza, Pastor-Satorras, and Vespignani, 2007)

$$\bar{\mathcal{N}} \geq \bar{\mathcal{N}}_c \equiv \frac{\langle k \rangle}{\langle k^2 \rangle} \frac{\mu}{\beta}, \tag{87}$$

where $\bar{\mathcal{N}} = \sum P(k)\mathcal{N}_k = \mathcal{N}/N$ is the average number of individuals per node. This result implies that a stationary state with infectious individuals is possible only if the particle density average $\bar{\mathcal{N}}$ is larger than a specific critical threshold. However, the network topological fluctuations affect the critical value. In particular, in heavy-tailed networks with $\langle k^2 \rangle \to \infty$ we have that $\bar{\mathcal{N}}_c \to 0$, i.e., topological fluctuations induce a vanishing of the threshold in the limit of an infinite network.

The different behavior obtained in the two types of processes can be understood qualitatively by the following argument (Colizza, Pastor-Satorras, and Vespignani, 2007). In a process governed by the mass-action law the epidemic activity in each node is rescaled by the local population $\mathcal{N}_i$ and it is therefore the same in all nodes. In this case, the generation of infected individuals is homogeneous across the network and an epidemic active state depends only on the balance between β and μ, whose values must poise the system above the critical threshold. In contagion processes determined by the pseudomass-action law, whatever the parameters β and μ, there exists a local density of individuals able to sustain the generation of infected individuals to keep the system in the active state. In this case topological fluctuations induce density fluctuations in the network as the diffusion process brings individuals to each node proportionally to the degree k, Eq. (82). Whatever the average number of individuals per node in the thermodynamic limit, there is always a node (with a virtually infinite degree) with enough individuals to keep alive the contagion process, leading to the disappearance of the phase transition.

While the above results are obtained by a discrete formulation that generally well suits simulation schemes in which reactions and diffusion are executed sequentially, the continuum formalism of these models was derived by Saldaña (2008) [see also Baronchelli, Catanzaro, and Pastor-Satorras (2008)]. In the continuum derivation the same phenomenology is obtained although the results concerning the critical value in pseudomass reactionlike processes scales as the maximum degree in the network $\bar{\mathcal{N}}_c \sim k_{\max}^{-1}$.

It is worth stressing that in most contagion processes the mobility of individuals is generally extremely heterogeneous and not simply mimicked by constant diffusion probabilities as those used in the previous simple example. The interaction among subpopulations is the result of the movement of individuals from one subpopulation to the other. For instance, it is clear that one of the key issues in the modeling of contagion phenomena in human populations is the accurate description of 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. More realistic descriptions are provided by approaches which include explicitly the detailed rate of traveling or commuting obtained from data or from an empirical fit to gravity law models (Viboud *et al.*, 2006). For analytical studies, simplified approaches use the Markovian assumption in which at each time step the movement of individuals is given according to a matrix d_{ij} that expresses the rate at which an individual in the subpopulation i is traveling to the subpopulation j. This approach is extensively used in large populations where the traffic w_{ij} between subpopulations is known, stating that $d_{ij} \sim w_{ij}/\mathcal{N}_j$. Several modeling approaches to the large-scale spreading of infectious disease (Baroyan *et al.*, 1969; Rvachev and Longini, 1985; Flahault and Valleron, 1992; Grais *et al.*, 2004; Hufnagel, Brockmann, and Geisel, 2004; Colizza *et al.*, 2006; Colizza, Barrat *et al.*, 2007; Balcan, Colizza *et al.*, 2009) use this mobility process based on real data about transportation networks. A detailed description of different mobility and diffusion schemes can be found in Colizza and Vespignani (2008).

B. SIR model in metapopulation networks and the global invasion threshold

In the analysis of contagion processes in metapopulation networks, the diffusion parameters that mimic the mobility rate of individuals or particles in the system may cause severe changes to the phase diagram by inducing a novel type of critical threshold. To see these effects we consider SIR-like models with no stationary state possible. If we assume a diffusion probability p for each individual and that the single population reproduction number of the SIR model is $R_0 > 1$, we can easily identify two different limits. If $p = 0$, any epidemic occurring in a given subpopulation will remain confined; no individual can travel to a different subpopulation and spread the infection across the system. In the limit $p \to 1$ we see that individuals are constantly wandering from one subpopulation to another and the system is in practice equivalent to a well-mixed unique population. In this case, since $R_0 > 1$, the epidemic will spread across the entire system. A transition point between these two regimes is therefore occurring at a threshold value p_c of the diffusion rate, identifying a global invasion threshold that depends on the mobility as well as the parameters of the contagion process (see Fig. 11). In other words, in a model such as the SIR model, the epidemic within each subpopulation generates a finite fraction of infectious individuals in a finite amount of time, and even if $R_0 > 1$ the diffusion rate must be large

enough to ensure the diffusion of infected individuals to other subpopulations before the local epidemic outbreak dies out. It is worth remarking that this does not apply in models with endemic states such as the SIS model. In this case the disease produces infectious individuals indefinitely in time and sooner or later the epidemic will be exported to other subpopulations.

The invasion threshold is encoded in a new quantity R_* characterizing the disease invasion of the metapopulation system. R_* denotes the number of subpopulations that become infected from a single initially infected subpopulation, i.e., the analog of the reproduction number R_0 at the subpopulation level. It defines the critical values of parameters that allow the contagion process to spread across a macroscopic fraction of subpopulations. Interestingly, this effect cannot be captured by a continuous description that would allow any fraction $p\bar{I}$ of diffusing infected individual to inoculate the virus in a subpopulation not yet infected. In certain conditions this fraction $p\bar{I}$, that is a mean-field average value, may be a number smaller than 1. This is a common feature of continuous approximations that allow the infection to persist and diffuse via "nanoindividuals" that are not capturing the discrete nature of the real systems. The discrete nature of individuals and the stochastic nature of the diffusion can therefore have a crucial role in the problem of resurgent epidemics, extinction, and eradication (Ball, Mollison, and Scalia-Tomba, 1997; Cross *et al.*, 2005, 2007; Watts *et al.*, 2005; Vazquez, 2007).

In order to provide an analytical estimate of the invasion threshold, we considered a metapopulation network with arbitrary degree distribution $P(k)$, where each node of degree k has a stationary population $\mathcal{N}_k$. By using a Levins-type approach (Colizza and Vespignani, 2007) it is possible to characterize the invasion dynamics by looking at the treelike branching process describing the contagion process at the subpopulation level (Levins, 1970). We define D_k^0 as the number of diseased subpopulations of degree k at generation 0, i.e., those which are experiencing an outbreak at the beginning of the process. Each infected subpopulation will seed—during the course of the outbreak—the infection in neighboring subpopulations, defining the set D_k^1 of infected subpopulations at generation 1, and so on. This corresponds to a basic branching process where the number of infected subpopulations of degree k at the nth generation is denoted as D_k^n. We can write the iterative equation relating D_k^n and D_k^{n-1} as

$$D_k^n = \sum_{k'} D_{k'}^{n-1}(k'-1)P(k|k')\left(1-\frac{D_k^{n-1}}{N_k}\right)\left[1-\left(\frac{1}{R_0}\right)^{\lambda_{k'k}}\right]. \tag{88}$$

In this expression we assume that each infected subpopulation of degree k' at the $(n-1)$th generation may seed the infection in a number of subpopulations of degree k according to the number of neighboring subpopulations $(k'-1)$ that discount the neighboring population from which the infection was originally transmitted. The right term takes into account the probability $P(k|k')$ that each of the $k'-1$ neighboring populations has degree k, the probability that the seeded population is not infected, and the probability to observe an outbreak in the seeded population. This last probability stems from the probability of extinction $P_{\text{ext}} = 1/R_0$ of an epidemic seeded with a single infectious individual (Bailey, 1975), when one considers a seed of size $\lambda_{kk'}$ given by the number of infected individuals that move into a connected subpopulation of degree k' during the duration of the local outbreak in the subpopulation of degree k.

The quantity $\lambda_{kk'}$ can be explicitly calculated by considering that, in the case of a macroscopic outbreak in a closed population, the total number of infected individuals during the outbreak evolution will be equal to $\bar{\alpha}\mathcal{N}_k$ where $\bar{\alpha}$ depends on the specific disease model and parameter values used. Each infected individual stays in the infectious state for a time μ^{-1} equal to the inverse of the recovery rate, during which it can travel to the neighboring subpopulation of degree k' with rate p. Here for simplicity we consider that the mobility coefficient p is the same for all individuals. Under this condition the number of infected individuals that may move into a connected subpopulation of degree k' during the duration of the local outbreak in the subpopulation of degree k is given by

$$\lambda_{kk'} = p\frac{\bar{\mathcal{N}}\,\bar{\alpha}\,\mu^{-1}}{\langle k\rangle}, \tag{89}$$

where we considered that each individual will diffuse with the same probability in any of the k available connections and that $\mathcal{N}_k$ is given by Eq. (82).

In order to provide an explicit solution to the above iterative equation we consider in the following that $R_0 - 1 \ll 1$, thus assuming that the system is very close to the epidemic threshold. In this limit we can approximate the outbreak probability as $1 - R_0^{-\lambda_{k'k}} \simeq \lambda_{k'k}(R_0-1)$. In addition, we assume that at the early stage of the epidemic $D_k^{n-1}/N_k \ll 1$, and we consider the case of uncorrelated networks, obtaining

$$D_k^n = (R_0-1)\frac{kP(k)}{\langle k\rangle^2}\frac{p\bar{\mathcal{N}}\,\bar{\alpha}}{\mu}\sum_{k'} D_{k'}^{n-1}(k'-1). \tag{90}$$

By defining $\Theta^n = \sum_{k'} D_{k'}^n(k'-1)$, the last expression can be conveniently written in the iterative form

$$\Theta^n = (R_0-1)\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle^2}\frac{p\bar{\mathcal{N}}\,\bar{\alpha}}{\mu}\Theta^{n-1} \tag{91}$$

that allows a growing epidemic only if

$$R_* = (R_0-1)\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle^2}\frac{p\bar{\mathcal{N}}\,\bar{\alpha}}{\mu} > 1, \tag{92}$$

defining the *global invasion threshold* of the metapopulation system.

The explicit form of the threshold condition can be used to find the minimum mobility rate ensuring that on average each subpopulation can seed more than one neighboring subpopulation. The constant $\bar{\alpha}$ is larger than zero for any $R_0 > 1$, and in the SIR case for R_0 close to 1 it can be approximated by

$\bar{\alpha} \simeq 2(R_0 - 1)/R_0^2$ (Bailey, 1975), yielding a critical mobility value p_c below which the epidemics cannot invade the metapopulation system given by

$$p_c \bar{\mathcal{N}} \geq \frac{\langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle} \frac{\mu R_0^2}{2(R_0 - 1)^2}. \tag{93}$$

In Fig. 12 we show the total fraction of infected individuals across all subpopulations, also called the global attack rate, as a function of both R_0 and p, as obtained from extensive Monte Carlo simulations in an uncorrelated metapopulation network with $P(k) \sim k^{-2.1}$, $N = 10^5$, $\bar{\mathcal{N}} = 10^3$, and $\mu = 0.2$. The global attack rate surface in the p-R_0 space shows that the smaller the value of R_0, the higher the mobility p in order for the contagion process to successfully invade a finite fraction of the subpopulations.

The invasion threshold $R_* > 1$ implicitly defines the critical mobility rate of individuals and is an indicator as important as the basic reproduction number $R_0 > 1$ in assessing the behavior of contagion processes in structured populations. It shifts the attention from the local outbreak to a global perspective where the interconnectivity and mobility among subpopulations is important in possibly hampering the spreading process. The presence of the factor $\langle k \rangle^2 / \langle k^2 \rangle$ in the explicit expression of the threshold points out that also at the global level the heterogeneity of the network plays an important role. In other words, the topological fluctuations favor the subpopulation invasion and suppress the phase transition in the infinite size limit.

While the analysis we presented here is simplified, in the last years several studies have provided insight on metapopulation spreading fully considering the stochastic and discrete nature of the process in various realistic contexts: heterogenous schemes for the diffusion of individuals (Colizza and Vespignani, 2008; Gautreau, Barrat, and Barthélemy, 2008; Ni and Weng, 2009; Ben-Zion, Cohen, and Shnerb, 2010), heterogeneous populations (Poletto, Tizzoni, and Colizza, 2012; Apolloni, Poletto, and Colizza, 2013), non-Markovian recurrent mobility patterns mimicking commuting among geographical regions (Balcan and Vespignani, 2011, 2012; Belik, Geisel, and Brockmann, 2011), and the introduction of individual behavioral responses to the presence of disease (Meloni *et al.*, 2011; Nicolaides, Cueto-Felgueroso, and Juanes, 2013). Indeed one of the interesting applications of the particle-network framework and the study of reaction-diffusion processes in metapopulation networks consists of providing analytic rationales for data-driven epidemic models.

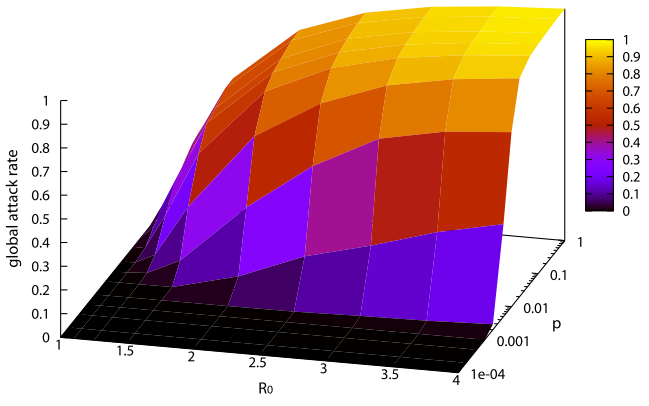

FIG. 12 (color online). Global threshold in a heterogeneous metapopulation system. The 3D surface representing the value of the final epidemic size in the metapopulation system as a function of the local threshold R_0 and of the diffusion probability p. If R_0 approaches the threshold, larger values of the diffusion probability p need to be considered in order to observe a global outbreak in the metapopulation system. Adapted from Colizza and Vespignani, 2007.

C. Agent based models and network epidemiology

In recent years, mathematical and computational approaches to the study of epidemics have been increasingly relevant in providing quantitative forecast and scenario analysis of real infectious disease outbreaks (Lofgren *et al.*, 2014). For this reason, epidemic models have evolved into large-scale microsimulations, data-driven approaches that can provide information at detailed spatial resolutions. An example is provided by agent based, spatially structured models that consider the discrete nature of individuals and their mobility and are generally including the stochasticity of interactions and the mobility of individuals. These models are based on the construction of synthetic populations characterizing each individual in the population and its mobility pattern, often down to the level of households, schools, and workplaces (Eubank *et al.*, 2004; Hufnagel, Brockmann, and Geisel, 2004; Ferguson *et al.*, 2005; Longini *et al.*, 2005; Colizza, Barrat *et al.*, 2007; Chao *et al.*, 2010). The synthetic population construction is a data hungry process and the resulting model is in most of the cases nontransparent to an analytical understanding. For this reason, the analysis of these models relies on computational microsimulations of the epidemic evolution that keep track of each single individual in the population. The resulting ensemble of possible epidemic evolutions is then leveraged to provide the usual quantitative indicators such as median, mean, and reference ranges for epidemic observables, such as newly generated cases, seeding events, and time of arrival of the infection. The statistical information generated by the computational approaches is then exploited with different visualization techniques that reference the data geographically. At first sight this modeling approach seems unrelated to network epidemiology. In reality, most of the data-driven computational approaches are relying on the construction of synthetic populations and interaction patterns that are effectively encoded as multiscale networks of individuals and locations (Marathe and Vullikanti, 2013).

An example of the underlying network structure of data-driven epidemic models is provided by the global epidemic and mobility (GLEAM) model that integrates census and mobility data in a fully stochastic metapopulation network model that allows for the detailed simulation of the spread of influenzalike illnesses around the globe (Broeck *et al.*, 2011). This model uses real demographic and mobility data. The world population is divided into geographic census areas that are defined around transportation hubs and connected by mobility fluxes. Within each subpopulation, the disease

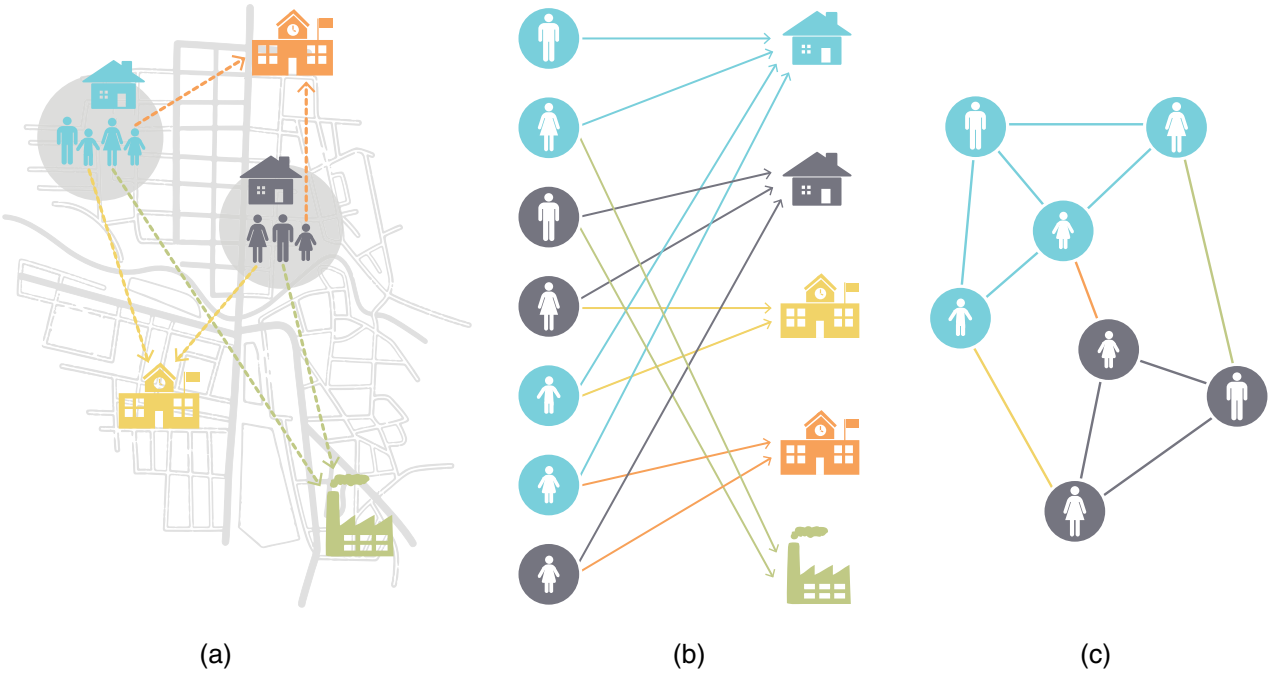

FIG. 13 (color online). Schematic illustration of the construction of a synthetic population and the resulting contact network. (a) At the macroscopic level, a synthetic population and its movements are constructed from census and demographic data. (b) A bipartite network associating individuals to locations, and eventually weighting the links with the time spent in the location, is derived from the synthetic population. (c) The unipartite projection of the bipartite network provides a contact network for the contagion process. Different transmission rates and weights on the network depend on the location and type of interactions.

spreads between individuals. Individuals can move from one subpopulation to another along the mobility network according to high quality transportation data, thus simulating the global spreading pattern of epidemic outbreaks. At the finer scale of urban areas, synthetic population constructions are even more refined and consider a classification of locations such as houses, schools, offices, etc. The movement and time spent in each location can be used to generate individual-location bipartite networks whose unipartite projection defines the individual-level, synthetic interaction network that governs the epidemic spreading (Eubank *et al.*, 2004; Halloran *et al.*, 2008; Merler *et al.*, 2011; Fumanelli *et al.*, 2012). Also in this case, although the model underlying the computational approach is a network model, each individual is annotated with the residence place and age, as well as other possible demographic information that can be exploited in the analysis of the epidemic outbreak (see Fig. 13).

Data-driven computational approaches can generate results at an unprecedented level of detail and have been used successfully in the analysis and forecast of real epidemics (Hufnagel, Brockmann, and Geisel, 2004; Balcan, Colizza *et al.*, 2009; Balcan *et al.*, 2009; Merler *et al.*, 2011) and a policy-making scenario analysis (Eubank *et al.*, 2004; Ferguson *et al.*, 2005; Longini *et al.*, 2005; Colizza, Barrat *et al.*, 2007; Brockmann and Helbing, 2013). Similar approaches are becoming more and more popular in the simulation of generalized contagion processes and social behavior (Marathe and Vullikanti, 2013). Although realistic and detailed computational approaches often provide non-intuitive results, the key mechanisms underlying the epidemic evolution are difficult to identify because of the amount of details integrated in the models. In such cases, the analytic understanding of the basic models presented in this review can be the key to the systematic investigation of the impact of the various complex features of real systems on the basic properties of epidemic outbreaks. For instance, the simple calculation of the invasion threshold explains why travel restrictions appear to be highly ineffective in containing epidemics in large-scale data-driven simulation: the complexity and heterogeneity of the present time human mobility network considerably favor the global spreading of infectious diseases. Only unfeasible mobility restrictions reducing the global travel fluxes by 90% or more would be effective (Cooper *et al.*, 2006; Hollingsworth, Ferguson, and Anderson, 2006; Colizza and Vespignani, 2008; Bajardi *et al.*, 2011). The understanding of the behavior of reaction-diffusion processes in complex networks is therefore a crucial undertaking if we want to answer many basic questions about the reliability and predictive power of data-driven computational models.

X. GENERALIZING EPIDEMIC MODELS AS SOCIAL CONTAGION PROCESSES

Infectious diseases certainly represent the central focus of epidemic modeling because of the relevance they played, and continue to play in present days, in human history. The contagion metaphor however applies in several other domains and, in particular, in the social context: the diffusion of information (Bikhchandani, Hirshleifer, and Welch, 1992), the propagation of rumors, and the adoption of innovations or behaviors (Bass, 1969; Rogers, 2010) are all phenomena for which the state of an individual is strongly influenced by the

interaction with peers. Mediated by the network of social contacts, these interactions can give rise to epidemiclike outbreaks: fads, information cascades, memes going viral online, etc. The term social (or complex) contagion generally denotes these types of phenomena. New communication technologies, online social media, and the abundance of digital fingerprints that we, as individuals, disseminate in our daily life provide an unprecedented wealth of data about social contagion phenomena, calling for theoretical approaches to measure, interpret, model, and predict them. Simple models for disease epidemics are the natural paradigm for this endeavor and have been applied to social spreading phenomena (Goffman and Newill, 1964; Goffman, 1966; Bettencourt *et al.*, 2006). Some specific features of social contagion, however, are qualitatively different from pathogen spreading: the transmission of information involves intentional acts by the sender and the receiver, it is often beneficial for both participants (as opposed to disease spreading), and it is influenced by psychological and cognitive factors. This leads to the introduction of new ingredients in the models, from which the name *complex contagion* derives. In this section we discuss recent developments in this modeling effort, which we divide into two broad categories depending on whether the spreading process (threshold models) or the recovery process (rumor spreading models) of the disease epidemic propagation is changed. In light of the modeling efforts, a review of papers analyzing empirical data follows next.

As the topics presented here encompass a vast spectrum of disciplines, including physics, computer science, mathematics, and social sciences, the usual caveat about the impossibility of an exhaustive review of all the literature is to be particularly stressed. Our limited goal is to try to outline the most important contributions in a unitary framework. This endeavor is made even more difficult by the fact that the propagation of social contagion is also close to other processes such as failure cascades [in network routing protocols or mechanical failure (Motter and Lai, 2002)] or the adoption of strategies in game-theoretic context (Easley and Kleinberg, 2010) that are beyond the scope of this review.

A. Threshold models

For disease epidemics it is customary to assume that a susceptible individual has a constant probability to receive the infection from a peer upon every exposure, independently of whether other infected individuals are simultaneously in contact or other exposures have occurred in the past. While generally reasonable for the transmission of pathogens [although exceptions may occur (Joh *et al.*, 2009)], this hypothesis is clearly unrealistic in most situations where a social meme is spreading: a piece of information is more credible if arriving from different sources; the push to adopt a technological innovation is stronger if neighboring nodes in the social network have already adopted it. These considerations lead naturally to the introduction of "threshold models" for spreading phenomena, where the effect of multiple exposures changes from low to high as a function of their number. Figure 14 displays the probability of infection (adoption) $P_{\rm inf}$ after K attempts in the different scenarios. In the case of SIR [Fig.14(a)] each attempt has a fixed probability p of success and $P_{\rm inf} = 1 - (1-p)^K$.

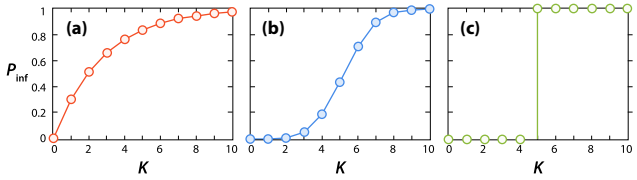

FIG. 14 (color online). Probability $P_{\rm inf}$ of infection for a susceptible individual after K contacts with infected individuals. (a) Independent interaction (e.g., SIR-type) model. (b) Stochastic threshold model. (c) Deterministic threshold model. Adapted from Dodds and Watts, 2004.

Threshold models have a long tradition in the social and economical sciences (Granovetter, 1978; Morris, 2000). In the context of spreading phenomena on complex networks, a seminal role has been played by the model introduced by Watts (2002). Each individual can be in one of two states (S and I) and is endowed with a quenched, randomly chosen threshold value ϕ_i. In an elementary step an individual agent in state S observes the current state of its neighbors and adopts state I if at least a threshold fraction ϕ_i of its neighbors is in state I; else it remains in state S.[8] No transition from I back to S is possible. Initially all nodes except for a small fraction are in state S. Out of these initiators a *cascade* of transitions to the I state is generated. The nontrivial question concerns whether the cascade remains local, i.e., restricted to a finite number of individuals, or it involves a finite fraction of the whole population. Given an initial seed, the spreading can occur only if at least one of its neighbors has a threshold such that $\phi_i \leq 1/k_i$. A cascade is possible only if a cluster of these "vulnerable" vertices is connected to the initiator. For global cascades to be possible it is then conjectured that the subnetwork of vulnerable vertices must percolate throughout the network. The condition for global cascades can then be derived applying on locally treelike networks the machinery of generating functions for branching processes. In the simple case of a uniform threshold ϕ and an Erdős-Rényi pattern of interactions the phase diagram as a function of the threshold ϕ and of the average degree $\langle k \rangle$ is reported in Fig. 15. For fixed ϕ, global cascades occur only for intermediate values of the mean connectivity $1 < \langle k \rangle < 1/\phi$. The transition occurring for small $\langle k \rangle$ is trivial and is not due to the spreading dynamics: the average cascade size is finite for $\langle k \rangle < 1$ because the network itself is composed of small disconnected components: the transition is percolative with power-law distributed cascade size. For large $\langle k \rangle > 1/\phi$ instead, the propagation is limited by the local stability of nodes. As the transition is approached increasing $\langle k \rangle$ the distribution of cascade size is bimodal, with an exponential tail at small cascade size and global cascades increasingly larger but more

[8]This is the definition for relative threshold models. In many cases absolute thresholds are considered (Granovetter, 1978; Kempe, Kleinberg, and Tardos, 2003; Centola and Macy, 2007; Galstyan and Cohen, 2007; Kimura *et al.*, 2009; Karimi and Holme, 2013). For strongly heterogeneous networks the different definitions may lead to important changes.

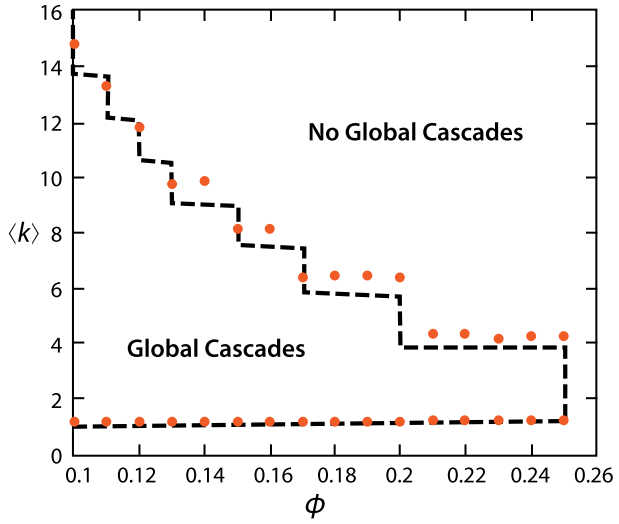

FIG. 15 (color online). Phase diagram of Watts's threshold model. The dashed line encloses the region of the $(\phi, \langle k \rangle)$ plane in which the condition for the existence of global cascades is satisfied for a uniform random graph with uniform threshold ϕ. The solid circles outline the region in which global cascades occur for the same parameter settings in the full dynamical model for $N = 10\,000$ (averaged over 100 random single-node perturbations). Adapted from Watts, 2002.

rare, until they disappear altogether, implying a discontinuous (i.e., first-order) phase transition in the size of successful cascades. Heterogeneous thresholds reduce the system stability, increasing the range of parameters where global cascades occur. Degree heterogeneity has instead the opposite effect.

The critical value of the threshold $\phi_c = 1/\langle k \rangle$ separating global cascades for $\phi < \phi_c$ from localized spreading for $\phi > \phi_c$ highlights the peculiar features of threshold dynamics (Centola, Eguíluz, and Macy, 2007). Adding new links to the network makes $\langle k \rangle$ grow, thus reducing ϕ_c and making system-wide spreading more difficult, the opposite of what occurs for SIR epidemics. Note indeed that the dependence of the threshold on the average degree is the same (for homogeneous networks) in both the threshold model and SIR dynamics, but in the latter case the global spreading occurs above the threshold (for $\lambda > 1/\langle k \rangle$), while in the former case global cascades are possible below the threshold ($\phi < 1/\langle k \rangle$). By the same token, link rewiring which destroys clustering of a network is seen to reduce the average cascade size for the threshold model. Instead of the strength of the weak ties (Granovetter, 1973) here the weakness of the long ties (Centola and Macy, 2007) is at work.

Watts's model can be seen as a particular instance of a more general model (Dodds and Watts, 2004), which includes also independent interaction models (SIR, SIRS) as particular cases. The model incorporates individual memory, a variable magnitude of exposure (dose amount), and heterogeneity in the susceptibility of individuals. At each contact with an infected neighbor a susceptible receives with probability p a random dose $d(t)$ [distributed according to $f(d)$]. A susceptible individual i accumulates the doses $d_i(t)$ over a time T and it becomes infected if at some time the accumulated dose $D_i(t) = \sum_{t'=t-T+1}^{t} d_i(t')$ is larger than a threshold d_i^* [random for each node with distribution $g(d^*)$]. Recovery is possible with probability r provided the dose $D_i(t)$ falls below d_i^*. The probability that a susceptible individual who encounters $K \leq T$ infected individuals in T time steps becomes infected is therefore

$$P_{\text{inf}}(K) = \sum_{k=1}^{K} \binom{K}{k} p^k (1-p)^{K-k} P_k, \tag{94}$$

where

$$P_k = \int_0^\infty dd^* g(d^*) P\left(\sum_{i=1}^{k} d_i \geq d^* \right) \tag{95}$$

is the average fraction of individuals infected after receiving k positive doses in T time steps. When all doses d_i are identical, all members of the population have the same threshold d^*, and $p < 1$, then the model reduces to the standard SIR [Fig. 14(a)]. In other cases it is a deterministic or stochastic threshold model, depending on whether thresholds vary [Fig. 14(b)] or are all identical [Fig. 14(c)].

Adding a probability ρ that a recovered individual becomes susceptible again leads to a SIRS-like dynamics. Setting $r = 1$ and $\rho = 1$ gives a SIS-like model, for which the stationary fraction of active nodes as a function of p is the order parameter. Three qualitatively different shapes of the phase diagram are found, depending only on T and P_1 and P_2, the probabilities that an individual will become infected as a result of one and two exposures, respectively. If $P_1 > P_2/2$ there is a standard epidemic transition between an absorbing healthy phase and an active infected one. The phenomenology is the same as SIS, indicating that successive exposures are effectively independent. The two other possible behaviors both exhibit a discontinuous phase transition for finite p, differing in the sensitivity with respect to the size of the initial seed.

By means of an analytical approach for locally treelike networks, Gleeson and Cahalane (2007) extended Watts's approach to consider a finite fraction of initiators p^{in}. It turns out that this change may have dramatic effects on the location of the transitions as a function of $\langle k \rangle$ and even make the transition for small $\langle k \rangle$ discontinuous. Singh *et al.* (2013) showed that for any $\phi < 1$ there is a critical value $p_c^{\text{in}}(\phi)$ such that for $p > p_c^{\text{in}}(\phi)$ the cascades are global. Further work along the same lines has generalized the analytical treatment to modular networks (Gleeson, 2008), degree-correlated networks (Gleeson, 2008; Dodds and Payne, 2009), and to networks with tunable clustering (Hackett, Melnik, and Gleeson, 2011). In the latter case, it turns out that for large and small values of $\langle k \rangle$ clustering reduces the size of cascades, while the converse occurs for intermediate values of the average degree.

Watts's threshold model has been extended in many directions, to take into account other potentially relevant effects that may influence the spreading process. Interaction patterns described by layered networks are found to increase the cascade size (Brummitt, Lee, and Goh, 2012) while the consideration of temporal networks (Holme and Saramäki, 2012) with the associated bursty activity of individuals may

either facilitate (Takaguchi, Masuda, and Holme, 2013) or hinder (Karimi and Holme, 2013) the spreading process. Watts's model on a basic two-community network is considered by Galstyan and Cohen (2007). Finally it is worth mentioning the work of Lorenz, Battiston, and Schweitzer (2009) who proposed a general classification of models for cascades, including, among many others, standard epidemic models and Watts's model as particular cases.

A large interest in threshold models has also been spurred by the goal of identifying influential spreaders, i.e., the starting nodes which maximize the size of cascades, a topic of interest also for traditional epidemic models (see Sec. VI.B). Kempe, Kleinberg, and Tardos (2003) showed that the problem of finding the set of initiator nodes such that the total size of the cascade is maximal (Domingos and Richardson, 2001) is nonpolynomial (NP) hard, for both linear threshold models and an independent cascade model, which is essentially an inhomogeneous SIR. Moreover, they provided a greedy hill-climbing algorithm that provides an efficient approximation to the NP-hard solution, outperforming random choice as well as choices based on degree centrality and distance centrality, when tested on some empirical networks. The method of Kempe, Kleinberg, and Tardos (2003) is computationally costly. An improvement which makes it much faster is provided by Kimura *et al.* (2009).

B. Rumor spreading

Models for rumor spreading are variants of the SIR model for disease epidemics in which the recovery process does not occur spontaneously, but rather is a consequence of interactions. The basic idea behind this modification is that it is worth propagating a rumor as long as it is novel for the recipient: If the spreader finds that the recipient already knows the rumor, he or she might lose interest in spreading it any further. The formalization of this process is due to Daley and Kendall (1964, 1965); individuals can be in one of three possible states[9]: ignorant (S, equivalent to susceptible in SIR), spreader (I, equivalent to infected), and stifler (R, equivalent to recovered). The possible events and the corresponding rates are

$$S + I \xrightarrow{\beta} 2I, \qquad R + I \xrightarrow{\alpha} 2R, \qquad 2I \xrightarrow{\alpha} 2R. \tag{96}$$

In a slightly distinct version, introduced by Maki and Thompson (1973), the third process is different: when a spreader contacts another agent and finds it in state I, only the former turns into a stifler, the latter remaining unchanged, i.e., the third process is

$$2I \xrightarrow{\alpha} R + I. \tag{97}$$

As for the SIR model, starting from a single informed individual the rumor propagates through the network with an increase in the number of spreaders. Asymptotically all spreaders turn into stiflers and in the final absorbing state there are only ignorants or stiflers. The "reliability," i.e., the fraction r_∞ of stiflers in this asymptotic state, quantifies whether the rumor remains localized ($r_\infty \to 0$ for system size $N \to \infty$) or spreads macroscopically. The solution of both versions of the model on the complete graph (Sudbury, 1985; Barrat, Barthélemy, and Vespignani, 2008) gives the whole temporal evolution of the reliability, yielding r_∞ as the solution of

$$r_\infty = 1 - e^{-(1+\beta/\alpha)r_\infty}. \tag{98}$$

As a consequence, r_∞ is positive for any $\beta/\alpha > 0$, i.e., the rumor spreads macroscopically for any value of the spreading parameters at odds with what happens for the SIR dynamics, which has a finite threshold for homogeneous networks.

Since models for disease epidemics are strongly affected by complex topologies, it is natural to ask what happens for rumor dynamics. When the Maki-Thompson model is simulated on scale-free networks it turns out that heterogeneity hinders the propagation dynamics by reducing the final reliability r_∞, still without introducing a finite threshold (Liu, Lai, and Ye, 2003; Moreno, Nekovee, and Pacheco, 2004; Moreno, Nekovee, and Vespignani, 2004). Why this happens is easily understood: large hubs are rapidly reached by the rumor, but then they easily turn into stiflers, thus preventing the further spreading of the rumor to their many other neighbors. This is confirmed by the observation that the density of ignorants of degree k at the end of the process decays exponentially with k (Moreno, Nekovee, and Vespignani, 2004). Degree-based mean-field approaches (Nekovee *et al.*, 2007; Zhou, Liu, and Li, 2007) are in good agreement with the numerical findings. The phenomenology of rumor spreading is markedly different from the behavior of the SIR model and this is due to the healing mechanism involving two individuals, present in both Maki-Thompson and Daley-Kendall dynamics. If spontaneous recovery is also allowed with rate μ, justified as the effect of forgetting, it turns out that the model behaves exactly as SIR: macroscopic spreading occurs only above a threshold inversely proportional to the second moment $\langle k^2 \rangle$, which then vanishes in the large network size limit for scale-free networks (Nekovee *et al.*, 2007). Again the interpretation of this outcome is not difficult: the forgetting term is linear in the density of spreaders and thus dominates for small densities, since the healing terms, due to the processes in Eqs. (96) and (97), are quadratic.

When the pattern of interactions among individuals is given by the Watts-Strogatz topology, rumor dynamics gives rise to a nontrivial phenomenon: a phase transition occurring at a critical value of the rewiring probability p (Zanette, 2001): For large values of p the network is essentially random and the rumor reaches a finite fraction of the vertices. For small values of p the spreading occurs only in a finite neighborhood of the initiator, so that the density of stiflers vanishes with the system size. In other transitions occurring on the Watts-Strogatz network, the critical point scales to zero with the system size N, a consequence of the fact that the geometric crossover between a one-dimensional lattice and a small-world structure scales as $1/N$ (Watts and Strogatz, 1998). Strikingly instead, the threshold p_c for macroscopic rumor spreading converges to a finite value as

[9]For consistency, we use the same symbols of the SIR model.

the system size grows. This indicates that the transition cannot be explained only in geometrical terms; some nontrivial interplay between topology and dynamics is at work. Interestingly, the transition at finite p_c persists also when an annealed Watts-Strogatz network is considered.

Recently, some activity has been devoted to the investigation of the role of influential spreaders in rumor spreading, in analogy to what has been done for disease epidemics (see Sec. VI.B). Borge-Holthoefer and Moreno (2012) looked for the role of nodes with large K-core index for the Maki-Thompson dynamics on several empirical networks. It turns out that the final density of stiflers does not depend on the K-core value of the initiator. Nodes with high-K-core index are not good spreaders; they are reached fast by the rumor and short-circuit its further spreading. An empirical investigation of cascades on the Twitter social network (Borge-Holthoefer, Rivero, and Moreno, 2012) points out instead that privileged spreaders (identified by large degree k or large K) do exist in real-world spreading phenomena, in patent contrast with the predictions of rumor spreading models. To reconcile theoretical predictions and empirical observations it is necessary to amend the Maki-Thompson dynamics. Two possible modifications have been proposed by Borge-Holthoefer *et al.* (2013). In one case individuals are not always active and do not spread further twits reaching them while inactive. In the second an ignorant contacted by a spreader turns into a spreader only with probability p, while with probability $1-p$ it turns directly into a stifler. The modified rumor spreading models are able to qualitatively reproduce the empirical findings, provided (for the first) that the probability to be active is proportional to the node degree or (for the second) that the probability p to actually spread is very small (of the order of 10^{-3}).

C. Empirical studies

Empirical data for a large number of spreading processes in the real world have been analyzed in terms of epidemiclike phenomena. Here we outline some of the most important contributions.

Leskovec, Adamic, and Huberman (2007) analyzed an instance of viral marketing, in the form of the email recommendation network for products of a large retailer. There are large variations depending on the type of goods recommended, its price, and the community of customers targeted, but in general recommendations turn out to be not very effective and cascades of purchases are not very extended. The key factor, different from disease epidemics, is that the "infection probability" quickly saturates to a low value with the number of recommendations received. Moreover, as an individual sends more and more recommendations the success per recommendation declines (high degree individuals are not so influent). Overall, viral marketing is very different from epidemiclike spreading.

A case where cascades are large and the spreading is a real collective phenomenon is the propagation of chain letters on the Internet. Liben-Nowell and Kleinberg (2008) found tree-like dissemination patterns, very deep but not large. A simple epidemiclike model, with an individual having a probability to forward the message to a fraction of his or her contacts, gives instead wide and shallow trees. More realistic propagations are obtained introducing two additional ingredients, asynchronous response times and "backresponse" (Liben-Nowell and Kleinberg, 2008).

Cascading behavior in large blog graphs is actively investigated (Gruhl *et al.*, 2004; Adar and Adamic, 2005). Leskovec *et al.* (2007) found that in this case cascades tend to be wide, not deep, with a size distribution following a power law with a slope of -2. The shape of cascades is often starlike. A single-parameter generative model (essentially a SIS-like model in the absorbing phase) is in good agreement with empirical observations regarding frequent cascade shapes and size distributions.

Also the behavior of individuals is subject to social influence and thus gives rise to collective spreading. Obesity, smoking habits, and even happiness (Christakis and Fowler, 2007, 2008; Fowler and Christakis, 2008) have been claimed to spread as epidemics in social networks [see, however, Shalizi and Thomas (2011) for a criticism of these results]. In an empirical investigation Centola (2010) analyzed an artificially structured online community, devised to check whether spreading is favored by random unclustered structures [as in the "strength of weak ties" hypothesis (Granovetter, 1973)] or by clustered ones with larger diameter (Centola and Macy, 2007). The latter structures turn out to favor spreading, the more so for increasing degree. At the individual level, the presence of two or three neighbors adopting a behavior leads to an increase in the probability of doing the same. For four and more neighbors the probability remains instead constant.

For a long time empirical investigations of spreading phenomena suffered from the drawback that the network mediating the propagation was unknown and its properties had to be in some way guessed from how the spreading process itself unfolds. Online social networks (such as Facebook and Twitter) are an ideal tool to bypass this problem as they provide both the topology of existing connections and the actual path followed by the spreading process on top of the contact graph (Lerman and Ghosh, 2010). In one of such social networks (Digg), Ver Steeg, Ghosh, and Lerman (2011) found that while the network of contacts has a scale-free degree distribution, the size of cascades is log-normally distributed, with essentially all propagations limited to a fraction smaller than 1% of the whole network. Within the framework of a SIR model this would imply that the spreading parameter of each cascade is fine-tuned around the transition point. Two additional ingredients help to reconcile the empirical findings with models: on the one hand, Digg contact network has a high clustering and this feature leads to a reduction of outbreak size; on the other hand, as in Centola (2010), the probability to transmit the spreading quickly saturates as a function of the number of active neighbors. Another empirical investigation of Digg (Doerr *et al.*, 2012) [see also Van Mieghem, Blenn, and Doerr (2011)] found that links between friends in the social network contribute surprisingly little to the propagation of information.

Another critical element of the spreading of memes in modern online social networks is the competition among a large number of them. Weng *et al.* (2013) analyzed Twitter, finding a broad variability of the lifetime and popularity of

spreading memes. A minimalistic model, based on the heterogeneous structure of Twitter graphs of followers and on "limited attention," i.e., the survival of memes in agents' memory for only a finite time due to competition with others, is sufficient to reproduce the empirical findings. Surprisingly, it is not necessary to assume a variability in the intrinsic appeal of memes to explain the heterogeneous persistence and popularity of individual memes.

Another information spreading experiment was performed by Iribarren and Moro (2009), in which subscribers to an online newsletter in 11 European countries were offered a reward to recommend it via email. The recommendations were tracked at every step by means of viral propagation and it was thus possible to reconstruct the recommendation cascades originated by 7154 initiators. The topology of the observed cascades is essentially treelike, in agreement with the results of Liben-Nowell and Kleinberg (2008), and of very small size, suggesting again a behavior at or below a possible critical point. The heterogeneity of the viral spreading process was quantified by looking at the distribution of time elapsed between receiving an invitation email, and forwarding it to another individual. This distribution can be fitted to a long-tailed log-normal form. On the other hand, the average number of informed individuals forwarding the message at time t was also found to decay slowly (with a log-normal shape), in contrast with the exponential decay expected in epidemics below the threshold. Similar results were reported for the retweet time of Twitter messages; see Doerr, Blenn, and Van Mieghem (2013).

XI. OUTLOOK

In the last years the entire field of epidemic modeling in networks has greatly progressed in the understanding of the interplay between network properties and contagion processes. We hope to have fairly portrayed the major advances and achieved clarity of presentation on the various theoretical and numerical approaches in a field that has literally exploded. However, the results and understanding achieved so far have opened the door to new questions and problems, often stimulated by the availability of new data. For this reason, the research activity has not slowed its pace and there is still a number of major challenges.

As shown in the previous sections, we are just moving the first steps to access the mathematical and statistical laws that characterize the coevolution mechanisms between the network evolution and the dynamical process. This is a key element in most social networks, where it is almost impossible to disentangle the agents cognitive processes shaping the network evolution and their perception and awareness of the contagion processes.

Indeed, the adaptive behavior of individuals in response to the dynamical processes they are involved in represents a serious theoretical challenge dealing with the feedback among different and competing dynamical processes. For instance, some activity has already been devoted to coupled behavior-disease models and to the competition among different contagion processes in networks, but much more work is needed to build a comprehensive picture. The final goal is not only to understand epidemic processes and predict their behavior, but also to control their dynamics. The development of strategies for favoring or hindering contagion processes is crucial in a wide range of applications that span from the optimization of disease containment and eradication to viral marketing. Also in this case, much more work is needed investigating how coevolution and feedback mechanisms between the network evolution and the spreading dynamics affect our influence and ability to control epidemic processes.

Networks show also a large number of interdependencies of various nature: physical interdependency when energy, material, or people flow from one infrastructure to another; cyber interdependency when information is transmitted or exchanged; geographic interdependency signaling the colocation of infrastructural elements; or logical interdependency due to financial, political coordination, etc. Interdependence is a major issue also in diffusion and spreading processes. One simple example is provided by the spreading of information in communication networks that induces an alteration of the physical proximity contact pattern of individuals or of the flows and traffic of mobility infrastructure. This has triggered interest in the understanding of contagion processes in coupled interdependent networks (Son *et al.*, 2012). More broadly, the community is becoming aware that, especially in the area of modern social networks populating the information technology ecosystem, epidemic spreading may occur on different interacting networks that however affect each other. This is the case of information processes where different types of social communication networks (phone, real-world, digital) coexist and contribute to the spreading process. This evidence has led recently to the introduction of multilayer or multiplex networks (Boccaletti *et al.*, 2014; Kivelä *et al.*, 2014). Multiplex networks are defined by a set of N nodes and a set of L layers, which represent "dimensions" or "aspects" that characterize a node. A node can belong to any subset of layers, and edges represent interactions between nodes belonging to the same layer. We can consider that a vertex is connected to itself across the different layers, or allow for interlayer connections between nodes in different layers. Every layer is represented thus by a network, and the whole multiplex by a set of interconnected networks. The analysis of epidemic processes in these networks shows interesting and peculiar behaviors. Several studies have focused on physical-information layered networks and studied the epidemic dynamics on the different layers as a function of the interlayer coupling and the epidemic threshold values on each layer (Marceau *et al.*, 2011; Yagan *et al.*, 2013; Buono *et al.*, 2014) For the SIR model it is also observed that depending on the average degree of interlayer connections (Dickison, Havlin, and Stanley, 2012) a global endemic state may arise in the interconnected system even if no epidemics can survive in each network separately (Saumell-Mendiola, Serrano, and Boguñá, 2012; Sahneh, Scoglio, and Chowdhury, 2013). SIS dynamics on multiple coupled layers has also been analyzed by Cozzo *et al.* (2013) and Sahneh, Scoglio, and Chowdhury (2013) in a generalized mean-field framework. However, epidemic behavior on multiplex networks is still largely unexplored for more complex models, complex contagion phenomena, and in data-driven settings.

The ever increasing computational power is also favoring detailed models that simulate large-scale population networks,

including geographic and demographic attributes on an individual by individual basis. These models can generate information at an unprecedented level of detail and guide researchers in identifying typical nonlinear behavior and critical points that often challenge our intuition. These results call for a theoretical understanding and a systematic classification of the models' dynamical behaviors, thus adding transparency to the numerical results. Results raise new general questions such as the following: What are the fundamental limits in the predictability of epidemics on networks? How does our understanding depend on the level of data aggregation and detail? What is the impact of the knowledge on the state and initial conditions of the network on our understanding of its dynamical behavior? These are all major conceptual and technical challenges that require the involvement of a vast research community and a truly interdisciplinary approach, rooted in the combination of large-scale data mining techniques, computational methods, and analytical techniques.

The study of epidemic spreading is a vibrant research area that is finding more and more applications in a wide range of domains. The need for quantitative and mathematical tools able to provide understanding in areas ranging from infectious diseases to viral marketing is fostering the intense research activity at the forefront in the investigation of epidemic spreading in networks. We hope that the present review will be a valuable reference for all researchers that will engage in this field.

ACKNOWLEDGMENTS

R. P.-S. acknowledges financial support from the Spanish MINECO, under Project No. FIS2010-21781-C02-01 and No. FIS2013-47282-C2-2, EC FET-Proactive Project MULTIPLEX (Grant No. 317532), and ICREA Academia, funded by the Generalitat de Catalunya. P. V. M. was partially funded by the EU CONGAS project (No. 288021). A. V. was partially funded by the DTRA-1-0910039, NSF CMMI-1125095, MIDAS-National Institute of General Medical Sciences U54GM111274 awards. We thank Nicole Samay for help with the diagrams and figures. We thank A. Barrat and J. Miller for the many useful suggestions at various stages of the manuscript preparation.

REFERENCES

Abramson, G., and M. Kuperman, 2001, Phys. Rev. Lett. **86**, 2909.
Adar, E., and L. Adamic, 2005, in *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, WI '05* (IEEE Computer Society, Washington, DC), pp. 207–214.
Ahn, Y.-Y., H. Jeong, N. Masuda, and J. D. Noh, 2006, Phys. Rev. E **74**, 066113.
Ahnert, S. E., D. Garlaschelli, T. M. A. Fink, and G. Caldarelli, 2007, Phys. Rev. E **76**, 016101.
Albert, R., and A.-L. Barabási, 2002, Rev. Mod. Phys. **74**, 47.
Altarelli, F., A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina, 2014, Phys. Rev. Lett. **112**, 118701.
Ancel, L. W., M. E. Newman, M. Martin, and S. Schrag, 2003, Emerging Infect. Dis. **9**, 204.
Anderson, R. M., and R. M. May, 1992, *Infectious Diseases in Humans* (Oxford University Press, Oxford).
Andersson, H., and T. Britton, 2000, *Stochastic Epidemic Models and Their Statistical Analysis*, Lecture Notes in Statistics, Vol. 151 (Springer, New York).
Apolloni, A., C. Poletto, and V. Colizza, 2013, BMC Infect. Dis. **13**, 176.
Bagnoli, F., P. Liò, and L. Sguanci, 2007, Phys. Rev. E **76**, 061904.
Bailey, N. T. J., 1975, *The Mathematical Theory of Infectious Diseases and its Applications* (Charlin Griffin & Company, London), 2nd ed.
Bajardi, P., C. Poletto, J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani, 2011, PLoS One **6**, e16591.
Balcan, D., V. Colizza, B. Goncalves, H. Hu, J. Ramasco, and A. Vespignani, 2009, Proc. Natl. Acad. Sci. U.S.A. **106**, 21484.
Balcan, D., and A. Vespignani, 2011, Nat. Phys. **7**, 581.
Balcan, D., and A. Vespignani, 2012, J. Theor. Biol. **293**, 87.
Balcan, D., *et al.*, 2009, BMC Med. **7**, 45.
Ball, F., D. Mollison, and G. Scalia-Tomba, 1997, Ann. Appl. Probab. **7**, 46.
Bancal, J.-D., and R. Pastor-Satorras, 2010, Eur. Phys. J. B **76**, 109.
Barabási, A.-L., 2015, *Network Science* (Cambridge University Press, Cambridge, England).
Barabási, A.-L., and R. Albert, 1999, Science **286**, 509.
Baronchelli, A., M. Catanzaro, and R. Pastor-Satorras, 2008, Phys. Rev. E **78**, 016111.
Baronchelli, A., R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, 2013, Trends Cogn. Sci. **17**, 348.
Baronchelli, A., and R. Pastor-Satorras, 2010, Phys. Rev. E **82**, 011111.
Baroyan, O. V., L. A. Genchikov, L. A. Rvachev, and V. A. Shashkov, 1969, Bull. Int. Epidemiol. Assoc. **18**, 22.
Barrat, A., M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, 2004, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747.
Barrat, A., M. Barthélemy, and A. Vespignani, 2004, Phys. Rev. Lett. **92**, 228701.
Barrat, A., M. Barthélemy, and A. Vespignani, 2008, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England).
Barthélemy, M., A. Barrat, R. Pastor-Satorras, and A. Vespignani, 2004, Phys. Rev. Lett. **92**, 178701.
Barthélemy, M., A. Barrat, R. Pastor-Satorras, and A. Vespignani, 2005, J. Theor. Biol. **235**, 275.
Bass, F. M., 1969, Management Science **15**, 215.
Bauch, C. T., and A. P. Galvani, 2013, Science **342**, 47.
Bauer, F., and J. T. Lizier, 2012, Europhys. Lett. **99**, 68007.
Belik, V., T. Geisel, and D. Brockmann, 2011, Phys. Rev. X **1**, 011001.
ben-Avraham, D., and S. Havlin, 2005, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, England).
ben-Avraham, D., and J. Köhler, 1992, Phys. Rev. A **45**, 8358.
Bender, E. A., and E. R. Canfield, 1978, J. Comb. Theory Ser. A **24**, 296.
Ben-Zion, Y., Y. Cohen, and N. M. Shnerb, 2010, J. Theor. Biol. **264**, 197.
Bernoulli, D., 1760, Histoire de l'Acad. Roy. Sci. (Paris) avec M. des Math. et Phys., pp. 1–45.
Bettencourt, L. M., A. Cintrón-Arias, D. a. I. Kaiser, and C. Castillo-Chávez, 2006, Physica (Amsterdam) **364A**, 513.
Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, J. Polit. Econ. **100**, 992.
Binder, K., and D. W. Heermann, 2010, *Monte Carlo Simulation in Statistical Physics* (Springer-Verlag, Berlin), 5th ed.
Bisanzio, D., L. Bertolotti, L. Tomassone, G. Amore, C. Ragagli, A. Mannelli, M. Giacobini, and P. Provero, 2010, PLoS One **5**, e13796.

Blythe, S. P., and R. M. Anderson, 1988, Math. Med. Biol. **5**, 181.
Boccaletti, S., G. Bianconi, R. Criado, C. del Genio, J. Gomez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, 2014, Phys. Rep. **544**, 1.
Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D. Hwang, 2006, Phys. Rep. **424**, 175.
Boguñá, M., C. Castellano, and R. Pastor-Satorras, 2009, Phys. Rev. E **79**, 036110.
Boguñá, M., C. Castellano, and R. Pastor-Satorras, 2013, Phys. Rev. Lett. **111**, 068701.
Boguñá, M., L. F. Lafuerza, R. Toral, and M. A. Serrano, 2014, Phys. Rev. E **90**, 042108.
Boguñá, M., and R. Pastor-Satorras, 2002, Phys. Rev. E **66**, 047104.
Boguñá, M., R. Pastor-Satorras, and A. Vespignani, 2003a, Phys. Rev. Lett. **90**, 028701.
Boguñá, M., R. Pastor-Satorras, and A. Vespignani, 2003b, in *Statistical Mechanics of Complex Networks*, Lecture Notes in Physics Vol. 625, edited by R. Pastor-Satorras, J. M. Rubí, and A. Díaz-Guilera (Springer-Verlag, Berlin), pp. 127–147.
Boguñá, M., R. Pastor-Satorras, and A. Vespignani, 2004, Eur. Phys. J. B **38**, 205.
Boguñá, M., and M. A. Serrano, 2005, Phys. Rev. E **72**, 016106.
Bonaccorsi, S., S. Ottaviano, F. De Pellegrini, A. Socievole, and P. Van Mieghem, 2014, Phys. Rev. E **90**, 012810.
Borge-Holthoefer, J., S. Meloni, B. Gonçalves, and Y. Moreno, 2013, J. Stat. Phys. **151**, 383.
Borge-Holthoefer, J., and Y. Moreno, 2012, Phys. Rev. E **85**, 026116.
Borge-Holthoefer, J., A. Rivero, and Y. Moreno, 2012, Phys. Rev. E **85**, 066123.
Brauer, F., and C. Castillo-Chavez, 2010, *Mathematical Models in Population Biology and Epidemiology*, Textsin Applied Mathematics Vol. 40 (Springer, New York), 2nd ed.
Braunstein, L. A., S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, 2003, Phys. Rev. Lett. **91**, 168701.
Britton, T., M. Deijfen, and F. Liljeros, 2011, J. Stat. Phys. **145**, 1368.
Brockmann, D., and D. Helbing, 2013, Science **342**, 1337.
Broeck, W., C. Gioannini, B. Goncalves, M. Quaggiotto, V. Colizza, and A. Vespignani, 2011, BMC Infect. Dis. **11**, 37.
Brummitt, C. D., K.-M. Lee, and K.-I. Goh, 2012, Phys. Rev. E **85**, 045102.
Buono, C., L. G. Alvarez-Zuzek, P. A. Macri, and L. A. Braunstein, 2014, PLoS One **9**, e92200.
Buono, C., F. Vazquez, P. A. Macri, and L. A. Braunstein, 2013, Phys. Rev. E **88**, 022813.
Butts, C. T., 2009, Science **325**, 414.
Caldarelli, G., 2007, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, Oxford).
Callaway, D. S., M. E. Newman, S. H. Strogatz, and D. J. Watts, 2000, Phys. Rev. Lett. **85**, 5468.
Castellano, C., and R. Pastor-Satorras, 2006, Phys. Rev. Lett. **96**, 038701.
Castellano, C., and R. Pastor-Satorras, 2010, Phys. Rev. Lett. **105**, 218701.
Castellano, C., and R. Pastor-Satorras, 2012, Sci. Rep. **2**, 371.
Catanzaro, M., M. Boguñá, and R. Pastor-Satorras, 2005, Phys. Rev. E **71**, 027103.
Cator, E., R. van de Bovenkamp, and P. Van Mieghem, 2013, Phys. Rev. E **87**, 062816.
Cator, E., and P. Van Mieghem, 2012, Phys. Rev. E **85**, 056111.
Cator, E., and P. Van Mieghem, 2013, Phys. Rev. E **87**, 012811.
Cattuto, C., W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, 2010, PLoS One **5**, e11596.
Centola, D., 2010, Science **329**, 1194.
Centola, D., V. M. Eguíluz, and M. W. Macy, 2007, Physica (Amsterdam) **374A**, 449.
Centola, D., and M. Macy, 2007, Am. J. Sociology **113**, 702.
Chakrabarti, D., Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, 2008, ACM Trans. on Information and System Security (TISSEC) **10**, 1.
Chao, D. L., M. E. Halloran, V. J. Obenchain, and I. M. Longini, Jr, 2010, PLoS Comput. Biol. **6**, e1000656.
Chatterjee, S., and R. Durrett, 2009, Ann. Probab. **37**, 2332.
Chen, D., L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, 2012, Physica (Amsterdam) **391A**, 1777.
Chen, D.-B., R. Xiao, A. Zeng, and Y.-C. Zhang, 2013, Europhys. Lett. **104**, 68006.
Chen, Y., G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, 2008, Phys. Rev. Lett. **101**, 058701.
Christakis, N. A., and J. H. Fowler, 2007, N. Engl. J. Med. **357**, 370.
Christakis, N. A., and J. H. Fowler, 2008, N. Engl. J. Med. **358**, 2249.
Christakis, N. A., and J. H. Fowler, 2010, PLoS One **5**, e12948.
Chu, X., Z. Zhang, J. Guan, and S. Zhou, 2011, Physica (Amsterdam) **390A**, 471.
Chung, F., L. Lu, and V. Vu, 2003, Proc. Natl. Acad. Sci. U.S.A. **100**, 6313.
Chung, N. N., L. Y. Chew, J. Zhou, and C. H. Lai, 2012, Europhys. Lett. **98**, 58004.
Cohen, R., D. ben-Avraham, and S. Havlin, 2002, Phys. Rev. E **66**, 036113.
Cohen, R., K. Erez, D. ben-Avraham, and S. Havlin, 2000, Phys. Rev. Lett. **85**, 4626.
Cohen, R., K. Erez, D. ben-Avraham, and S. Havlin, 2001, Phys. Rev. Lett. **86**, 3682.
Cohen, R., and S. Havlin, 2010, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, England).
Cohen, R., S. Havlin, and D. ben-Avraham, 2003, Phys. Rev. Lett. **91**, 247901.
Colizza, V., A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani, 2007, PLoS Med. **4**, e13.
Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani, 2006, Proc. Natl. Acad. Sci. U.S.A. **103**, 2015.
Colizza, V., R. Pastor-Satorras, and A. Vespignani, 2007, Nat. Phys. **3**, 276.
Colizza, V., and A. Vespignani, 2007, Phys. Rev. Lett. **99**, 148701.
Colizza, V., and A. Vespignani, 2008, J. Theor. Biol. **251**, 450.
Comin, C. H., and L. da Fontoura Costa, 2011, Phys. Rev. E **84**, 056105.
Cooper, B., R. Pitman, W. Edmunds, and N. Gay, 2006, PLoS Med. **3**, e212.
Costa, L. d. F., F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, 2007, Adv. Phys. **56**, 167.
Cox, D. R., 1967, *Renewal Theory* (Methuen, London).
Cozzo, E., R. A. Banos, S. Meloni, and Y. Moreno, 2013, Phys. Rev. E **88**, 050801.
Cross, P., P. Johnson, J. Lloyd-Smith, and M. Wayne, 2007, J. R. Soc. Interface **4**, 315.
Cross, P., J. Lloyd-Smith, P. Johnson, and M. Wayne, 2005, Ecol. Lett. **8**, 587.
Daley, D. J., and D. G. Kendall, 1964, Nature (London) **204**, 1118.
Daley, D. J., and D. G. Kendall, 1965, IMA J. Appl. Math. **1**, 42.
Danon, L., A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, 2011, Interdiscip. Perspect. Infect. Dis. **2011**, 284909.

da Silva, R. A. P., M. P. Viana, and L. da Fontoura Costa, **2012**, J. Stat. Mech. P07005.

Decreusefond, L., J.-S. Dhersin, P. Moyal, and V. C. Tran, 2012, Ann. Appl. Probab. **22**, 541.

Deijfen, M., 2011, Math. Biosci. **232**, 57.

de Oliveira, M. M., and R. Dickman, 2005, Phys. Rev. E **71**, 016129.

Dezsö, Z., and A.-L. Barabási, 2002, Phys. Rev. E **65**, 055103.

Dickison, M., S. Havlin, and H. E. Stanley, 2012, Phys. Rev. E **85**, 066109.

Diekmann, O., H. Heesterbeek, and T. Britton, 2012, *Mathematical Tools for Understanding Infectious Disease Dynamics* (Princeton University Press, Princeton, NJ).

Diekmann, O., and J. Heesterbeek, 2000, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (John Wiley & Sons, New York).

Dodds, P., and J. Payne, 2009, Phys. Rev. E **79**, 066115.

Dodds, P., and D. J. Watts, 2004, Phys. Rev. Lett. **92**, 218701.

Dodds, P. S., K. D. Harris, and J. L. Payne, 2011, Phys. Rev. E **83**, 056122.

Doerr, C., N. Blenn, S. Tang, and P. Van Mieghem, 2012, Comput. Commun. **35**, 796.

Doerr, C., N. Blenn, and P. Van Mieghem, 2013, PLoS One **8**, e64349.

Domingos, P., and M. Richardson, 2001, in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '01* (ACM, New York), pp. 57–66.

Dorogovtsev, S. N., 2010, *Lectures on Complex Networks*, Oxford Master Series in Physics (Oxford University Press, Oxford).

Dorogovtsev, S. N., A. V. Goltsev, and J. F. F. Mendes, 2008, Rev. Mod. Phys. **80**, 1275.

Dorogovtsev, S. N., and J. F. F. Mendes, 2002, Adv. Phys. **51**, 1079.

Dorogovtsev, S. N., and J. F. F. Mendes, 2003, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford).

Dorogovtsev, S. N., J. F. F. Mendes, and A. N. Samukhin, 2000, Phys. Rev. Lett. **85**, 4633.

Dorogovtsev, S. N., J. F. F. Mendes, and A. N. Samukhin, 2001, Phys. Rev. E **64**, 025101.

Dunbar, R. I., 1998, Evolutionary anthropology : issues, news, and reviews **6**, 178.

Durrett, R., 2010, Proc. Natl. Acad. Sci. U.S.A. **107**, 4491.

Eames, K. T., J. M. Read, and W. J. Edmunds, 2009, Epidemics **1**, 70.

Eames, K. T. D., and M. J. Keeling, 2002, Proc. Natl. Acad. Sci. U.S.A. **99**, 13330.

Easley, D., and J. Kleinberg, 2010, *Networks, Crowds, and Markets* (Cambridge University Press, Cambridge, England).

Eguíluz, V. M., and K. Klemm, 2002, Phys. Rev. Lett. **89**, 108701.

Erdős, P., and P. Rényi, 1959, Publ. Math. Debrecen **6**, 290.

Eubank, S., H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, 2004, Nature (London) **429**, 180.

Ferguson, N. M., D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke, 2005, Nature (London) **437**, 209.

Ferreira, S. C., C. Castellano, and R. Pastor-Satorras, 2012, Phys. Rev. E **86**, 041125.

Ferreira, S. C., R. S. Ferreira, C. Castellano, and R. Pastor-Satorras, 2011, Phys. Rev. E **84**, 066102.

Flahault, A., and A.-J. Valleron, 1992, Mathematical Population Studies **3**, 161.

Fortunato, S., 2010, Phys. Rep. **486**, 75.

Fowler, J. H., and N. A. Christakis, 2008, Br. Med. J. **337**, a2338.

Freeman, L. C., 1977, Sociometry **40**, 35.

Fujiwara, N., J. Kurths, and A. Díaz-Guilera, 2011, Phys. Rev. E **83**, 025101.

Fumanelli, L., M. Ajelli, P. Manfredi, A. Vespignani, and S. Merler, 2012, PLoS Comput. Biol. **8**, e1002673.

Funk, S., E. Gilad, C. Watkins, and V. A. A. Jansen, 2009, Proc. Natl. Acad. Sci. U.S.A. **106**, 6872.

Funk, S., and V. A. A. Jansen, 2010, Phys. Rev. E **81**, 036118.

Funk, S., M. Salathé, and V. a. a. Jansen, 2010, J. R. Soc. Interface **7**, 1247.

Gallos, L. K., and P. Argyrakis, 2004, Phys. Rev. Lett. **92**, 138301.

Galstyan, A., and P. Cohen, 2007, Phys. Rev. E **75**, 036109.

Ganesh, A., L. Massoulie, and D. Towsley, 2005, in *Proceedings of the IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 2 (IEEE, New York), pp. 1455–1466 [doi:10.1109/INFCOM.2005.1498374].

Gang, Y., Z. Tao, W. Jie, F. Zhong-Qian, and W. Bing-Hong, 2005, Chin. Phys. Lett. **22**, 510.

Gantmacher, F. R., 1974, *The Theory of Matrices*, Vol. II (Chelsea Publishing Company, New York).

Garas, A., F. Schweitzer, and S. Havlin, 2012, New J. Phys. **14**, 083030.

Garcia-Herranz, M., E. Moro, M. Cebrian, N. A. Christakis, and J. H. Fowler, 2014, PLoS One **9**, e92413.

Gautreau, A., A. Barrat, and M. Barthélemy, 2008, J. Theor. Biol. **251**, 509.

Gil, S., and D. Zanette, 2005, Eur. Phys. J. B **47**, 265.

Gilbert, E. N., 1959, Ann. Math. Stat. **30**, 1141.

Gillespie, D. T., 1977, J. Phys. Chem. **81**, 2340.

Givan, O., N. Schwartz, A. Cygelberg, and L. Stone, 2011, J. Theor. Biol. **288**, 21.

Gleeson, J., and D. Cahalane, 2007, Phys. Rev. E **75**, 056103.

Gleeson, J. P., 2008, Phys. Rev. E **77**, 046117.

Gleeson, J. P., 2011, Phys. Rev. Lett. **107**, 068701.

Gleeson, J. P., 2013, Phys. Rev. X **3**, 021004.

Goffman, W., 1966, Nature (London) **212**, 449.

Goffman, W., and V. A. Newill, 1964, Nature (London) **204**, 225.

Goltsev, A. V., S. N. Dorogovtsev, and J. F. F. Mendes, 2008, Phys. Rev. E **78**, 051105.

Goltsev, A. V., S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, 2012, Phys. Rev. Lett. **109**, 128702.

Gómez, S., A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, 2010, Europhys. Lett. **89**, 38009.

Gomez-Gardenes, J., P. Echenique, and Y. Moreno, 2006, Eur. Phys. J. B **49**, 259.

Gomez-Gardenes, J., V. Latora, Y. Moreno, and E. Profumo, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 1399.

Gonçalves, B., N. Perra, and A. Vespignani, 2011, PLoS One **6**, e22656.

Gourdin, E., J. Omic, and P. Van Mieghem, 2011, *8th International Workshop on Design of Reliable Communication Networks (DRCN 2011), Krakow, Poland* (IEEE, New York).

Grais, R., J. Eliis, A. Kress, and G. Glass, 2004, Health Care Manag. Sci. **7**, 127.

Granell, C., S. Gómez, and A. Arenas, 2013, Phys. Rev. Lett. **111**, 128701.

Granovetter, M., 1978, Am. J. Sociology **83**, 1420.

Granovetter, M. S., 1973, Am. J. Sociology **78**, 1360.

Grassberger, P., 1983, Math. Biosci. **63**, 157.

Grenfell, B., and J. Harwood, 1997, Trends in Ecology & Evolution **12**, 395.

Gross, T., and B. Blasius, 2008, J. R. Soc. Interface **5**, 259.

Gross, T., C. D'Lima, and B. Blasius, 2006, Phys. Rev. Lett. **96**, 208701.
Gruhl, D., R. Guha, D. Liben-Nowell, and A. Tomkins, 2004, in *Proceedings of the 13th Conference on World Wide Web-WWW '04* (ACM Press, New York), p. 491.
Guo, D., S. Trajanovski, R. van de Bovenkamp, H. Wang, and P. Van Mieghem, 2013, Phys. Rev. E **88**, 042802.
Hackett, A., S. Melnik, and J. Gleeson, 2011, Phys. Rev. E **83**, 056107.
Halloran, M. E., *et al.*, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 4639.
Hamilton, K. E., and L. P. Pryadko, 2014, Phys. Rev. Lett. **113**, 208701.
Hammersley, J. M., and D. Welsh, 1965, in *Bernoulli 1713, Bayes 1763, Laplace 1813* (Springer, New York), pp. 61–110.
Hanski, I., and O. Gaggiotti, 2004, *Ecology, Genetics and Evolution of Metapopulations* (Elsevier Science, Princeton).
Harris, T. E., 1974, Ann. Probab. **2**, 969.
Hébert-Dufresne, L., A. Allard, J.-g. Young, and L. J. Dubé, 2013, Sci. Rep. **3**, 2171.
Henkel, M., H. Hinrichsen, and S. Lübeck, 2008, *Non-equilibrium Phase Transition: Absorbing Phase Transitions* (Springer-Verlag, Netherlands).
Hernández, D. G., and S. Risau-Gusman, 2013, Phys. Rev. E **88**, 052801.
Hethcote, H., and J. Yorke, 1984, *Gonorrhea Transmission Dynamics and Control, Lecture Notes in Biomathematics* (Springer-Verlag, Berlin).
Hethcote, H. W., 2000, SIAM Rev. **42**, 599.
Hoffmann, T., M. A. Porter, and R. Lambiotte, 2012, Phys. Rev. E **86**, 046102.
Hollingsworth, T., N. Ferguson, and R. Anderson, 2006, Nat. Med. **12**, 497.
Holme, P., 2004, Europhys. Lett. **68**, 908.
Holme, P., 2005, Phys. Rev. E **71**, 046119.
Holme, P., B. J. Kim, C. N. Yoon, and S. K. Han, 2002, Phys. Rev. E **65**, 056109.
Holme, P., and J. Saramäki, 2012, Phys. Rep. **519**, 97.
Holme, P., and J. Saramäki, 2013, Eds., *Temporal Networks* (Springer, Berlin).
Hou, B., Y. Yao, and D. Liao, 2012, Physica (Amsterdam) **391A**, 4012.
House, T., and M. J. Keeling, 2010, J. R. Soc. Interface **8**, 67.
Huang, W., and C. Li, **2007**, J. Stat. Mech. P01014.
Hufnagel, L., D. Brockmann, and T. Geisel, 2004, Proc. Natl. Acad. Sci. U.S.A. **101**, 15124.
Hui, P., A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, 2005, in *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking* (ACM, New York), pp. 244–251.
Iribarren, J., and E. Moro, 2009, Phys. Rev. Lett. **103**, 038702.
Jackson, M., 2010, *Social and Economic Networks* (Princeton University Press, Princeton).
Janson, S., M. Luczak, and P. Windridge, 2014, Random Struct. Algorithms **45**, 726.
Jo, H.-H., J. I. Perotti, K. Kaski, and J. Kertész, 2014, Phys. Rev. X **4**, 011041.
Joh, R. I., H. Wang, H. Weiss, and J. S. Weitz, 2009, Bull. Math. Biol. **71**, 845.
Joo, J., and J. L. Lebowitz, 2004, Phys. Rev. E **69**, 066105.
Juhász, R., G. Ódor, C. Castellano, and M. A. Muñoz, 2012, Phys. Rev. E **85**, 066125.
Karimi, F., and P. Holme, 2013, Physica (Amsterdam) **392A**, 3476.
Karrer, B., and M. E. J. Newman, 2010, Phys. Rev. E **82**, 016101.
Karrer, B., and M. E. J. Newman, 2011, Phys. Rev. E **84**, 036106.
Karrer, B., M. E. J. Newman, and L. Zdeborová, 2014, Phys. Rev. Lett. **113**, 208702.
Karsai, M., R. Juhász, and F. Iglói, 2006, Phys. Rev. E **73**, 036116.
Karsai, M., M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, 2011, Phys. Rev. E **83**, 025102.
Ke, H., and T. Yi, 2006, Chin. Phys. **15**, 2782.
Keeling, M., and P. Rohani, 2007, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton).
Keeling, M. J., 1999, Proc. R. Soc. B **266**, 859.
Keeling, M. J., and K. T. D. Eames, 2005, J. R. Soc. Interface **2**, 295.
Keeling, M. J., and B. T. Grenfell, 1997, Science **275**, 65.
Keeling, M. J., and P. Rohani, 2002, Ecol. Lett. **5**, 20.
Kempe, D., J. Kleinberg, and E. Tardos, 2003, in *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York), pp. 137–146 [doi:10.1145/956750.956769].
Kenah, E., and J. C. Miller, 2011, Interdiscip. Perspect. Infect. Dis. **2011**, 543520.
Kenah, E., and J. M. Robins, 2007, Phys. Rev. E **76**, 036113.
Kermack, W. O., and A. G. McKendrick, 1927, Proc. R. Soc. A **115**, 700.
Kesten, H., 2003, in *From Classical to Modern Probability* (Springer, New York), pp. 93–143.
Kimura, M., K. Saito, R. Nakano, and H. Motoda, 2009, Data Mining and Knowledge Discovery **20**, 70.
Kiss, I. Z., C. G. Morris, F. Sélley, P. L. Simon, and R. R. Wilkinson, 2015, J. Math. Biol. **70**, 437.
Kitsak, M., L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, 2010, Nat. Phys. **6**, 888.
Kivelä, M., A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, 2014, J. Complex Networks **2**, 203.
Kivelä, M., R. K. Pan, K. Kaski, J. Kertész, J. Saramäki, and M. Karsai, **2012**, J. Stat. Mech. P03005.
Klemm, K., M. Á. Serrano, V. M. Eguíluz, and M. San Miguel, 2012, Sci. Rep. **2**, 292.
Kooij, R., P. Schumm, C. Scoglio, and M. Youssef, 2009, Networking 2009, **5550**, 562.
Krone, S. M., 1999, Ann. Appl. Probab. **9**, 331.
Lagorio, C., M. Dickison, F. Vazquez, L. A. Braunstein, P. A. Macri, M. V. Migueles, S. Havlin, and H. E. Stanley, 2011, Phys. Rev. E **83**, 026102.
Lagorio, C., M. Migueles, L. Braunstein, E. Lpez, and P. Macri, 2009, Physica (Amsterdam) **388A**, 755.
Lambiotte, R., L. Tabourier, and J.-C. Delvenne, 2013, Eur. Phys. J. B **86**, 320.
Lee, H. K., P.-S. Shim, and J. D. Noh, 2013, Phys. Rev. E **87**, 062812.
Lee, S., L. E. C. Rocha, F. Liljeros, and P. Holme, 2012, PLoS One **7**, e36439.
Lerman, K., and R. Ghosh, 2010, in *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)* (The AAAI Press, Menlo Park, CA), p. 90.
Leskovec, J., L. Adamic, and B. Huberman, 2007, ACM Transactions on the Web **1**, 5.
Leskovec, J., M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, 2007, "Patterns of cascading behavior in large blog graphs," in *SDM*, Vol. 7, pp. 551–556.
Levins, R., 1970, "*Extinction*," Lectures on Mathematics in the Life Sciences Vol. 2, pp. 75–107.
Li, C., R. van de Bovenkamp, and P. Van Mieghem, 2012, Phys. Rev. E **86**, 026116.
Li, C., H. Wang, and P. Van Mieghem, 2013, Phys. Rev. E **88**, 062802.

Li, P., J. Zhang, X.-K. Xu, and M. Small, 2012, Chin. Phys. Lett. **29**, 048903.

Liben-Nowell, D., and J. Kleinberg, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 4633.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, 2001, Nature (London) **411**, 907.

Lindquist, J., J. Ma, P. Driessche, and F. Willeboordse, 2011, J. Math. Biol. **62**, 143.

Liu, J., Y. Tang, and Z. Yang, **2004**, J. Stat. Mech., P08008.

Liu, J.-G., Z.-M. Ren, and Q. Guo, 2013, Physica (Amsterdam) **392A**, 4154.

Liu, S., N. Perra, M. Karsai, and A. Vespignani, 2014, Phys. Rev. Lett. **112**, 118702.

Liu, Z., and B. Hu, 2005, Europhys. Lett. **72**, 315.

Liu, Z., Y.-C. Lai, and N. Ye, 2003, Phys. Rev. E **67**, 031911.

Lloyd, A. L., 2001a, Proc. R. Soc. B **268**, 985.

Lloyd, A. L., 2001b, Theor. Popul. Biol. **60**, 59.

Lloyd, A. L., and R. M. May, 2001, Science **292**, 1316.

Lofgren, E. T., *et al.*, 2014, Proc. Natl. Acad. Sci. U.S.A. **111**, 18095.

Lokhov, A. Y., M. Mézard, H. Ohta, and L. Zdeborová, 2014, Phys. Rev. E **90**, 012801.

Longini, I. M., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. T. Cummings, and M. E. Halloran, 2005, Science **309**, 1083.

Lorenz, J., S. Battiston, and F. Schweitzer, 2009, Eur. Phys. J. B **71**, 441.

Ludwig, D., 1975, Math. Biosci. **23**, 33.

Maki, D. P., and M. Thompson, 1973, *Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences* (Prentice-Hall, Englewood Cliffs).

Marathe, M., and A. K. S. Vullikanti, 2013, Commun. ACM **56**, 88.

Marceau, V., P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, 2010, Phys. Rev. E **82**, 036116.

Marceau, V., P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, 2011, Phys. Rev. E **84**, 026105.

Marder, M., 2007, Phys. Rev. E **75**, 066103.

Marro, J., and R. Dickman, 1999, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, England).

Martin, T., X. Zhang, and M. E. J. Newman, 2014, Phys. Rev. E **90**, 052808.

Maslov, S., and K. Sneppen, 2002, Science **296**, 910.

Maslov, S., K. Sneppen, and A. Zaliznyak, 2004, Physica (Amsterdam) **333A**, 529.

Masuda, N., and N. Konno, 2006, J. Theor. Biol. **243**, 64.

Mata, A. S., and S. C. Ferreira, 2013, Europhys. Lett. **103**, 48003.

May, R. M., and A. L. Lloyd, 2001, Phys. Rev. E **64**, 066112.

Meloni, S., N. Perra, A. Arenas, S. Gomes, Y. Moreno, and A. Vespignani, 2011, Sci. Rep. **1**, 62.

Merler, S., M. Ajelli, A. Pugliese, and N. M. Ferguson, 2011, PLoS Comput. Biol. **7**, e1002205.

Meyers, L. A., M. Newman, and B. Pourbohloul, 2006, J. Theor. Biol. **240**, 400.

Miller, J., 2007, Phys. Rev. E **76**, 010101.

Miller, J., 2011, J. Math. Biol. **62**, 349.

Miller, J. C., 2009a, Phys. Rev. E **80**, 020901.

Miller, J. C., 2009b, J. R. Soc. Interface **6**, 1121.

Miller, J. C., 2013, Phys. Rev. E **87**, 060801.

Miller, J. C., A. C. Slim, and E. M. Volz, 2012, J. R. Soc. Interface **9**, 890.

Miller, J. C., and E. M. Volz, 2013, PLoS One **8**, e69162.

Min, B., K.-I. Goh, and I.-M. Kim, 2013, Europhys. Lett. **103**, 50002.

Min, B., K.-I. Goh, and A. Vazquez, 2011, Phys. Rev. E **83**, 036102.

Miritello, G., E. Moro, R. Lara, R. Martínez-López, J. Belchamber, S. G. Roberts, and R. I. Dunbar, 2013, Soc. Networks **35**, 89.

Molloy, M., and B. Reed, 1995, Random Struct. Algorithms **6**, 161.

Monasson, R., 1999, Eur. Phys. J. B **12**, 555.

Moore, C., and M. E. J. Newman, 2000, Phys. Rev. E **61**, 5678.

Moreno, Y., J. B. Gómez, and A. F. Pacheco, 2003, Phys. Rev. E **68**, 035103.

Moreno, Y., M. Nekovee, and A. F. Pacheco, 2004, Phys. Rev. E **69**, 066130.

Moreno, Y., M. Nekovee, and A. Vespignani, 2004, Phys. Rev. E **69**, 055101.

Moreno, Y., R. Pastor-Satorras, and A. Vespignani, 2002, Eur. Phys. J. B **26**, 521.

Morris, S., 2000, Rev. Economic Studies **67**, 57.

Motter, A., and Y.-C. Lai, 2002, Phys. Rev. E **66**, 065102.

Mountford, T., J.-C. Mourrat, D. Valesin, and Q. Yao, 2013, arXiv:1203.2972v1.

Muñoz, M. A., R. Juhász, C. Castellano, and G. Ódor, 2010, Phys. Rev. Lett. **105**, 128701.

Nardini, C., B. Kozma, and A. Barrat, 2008, Phys. Rev. Lett. **100**, 158701.

Nekovee, M., Y. Moreno, G. Bianconi, and M. Marsili, 2007, Physica (Amsterdam) **374A**, 457.

Newman, M., 2010, *Networks: An Introduction* (Oxford University Press, New York).

Newman, M. E. J., 2002a, Phys. Rev. Lett. **89**, 208701.

Newman, M. E. J., 2002b, Phys. Rev. E **66**, 016128.

Newman, M. E. J., 2003a, Phys. Rev. E **68**, 026121.

Newman, M. E. J., 2003b, SIAM Rev. **45**, 167.

Newman, M. E. J., 2005, Phys. Rev. Lett. **95**, 108701.

Newman, M. E. J., and C. R. Ferrario, 2013, PLoS One **8**, e71321.

Newman, M. E. J., S. H. Strogatz, and D. J. Watts, 2001, Phys. Rev. E **64**, 026118.

Ni, S., and W. Weng, 2009, Phys. Rev. E **79**, 016111.

Nian, F., and X. Wang, 2010, J. Theor. Biol. **264**, 77.

Nicolaides, C., L. Cueto-Felgueroso, and R. Juanes, 2013, J. R. Soc. Interface **10**, 0495.

Nishiura, N., 2011, European Journal of Epidemiology **26**, 583.

Noël, P.-A., A. Allard, L. u. Hébert-Dufresne, V. Marceau, and L. J. Dubé, 2012, Phys. Rev. E **85**, 031118.

Noël, P.-A., B. Davoudi, R. C. Brunham, L. J. Dubé, and B. Pourbohloul, 2009, Phys. Rev. E **79**, 026101.

Noh, J. D., and H. Rieger, 2004, Phys. Rev. Lett. **92**, 118701.

Nsoesie, E. O., J. S. Brownstein, N. Ramakrishnan, and M. V. Marathe, 2014, Influenza and Other Respiratory Viruses **8**, 309.

Ódor, G., 2013a, Phys. Rev. E **87**, 042132.

Ódor, G., 2013b, Phys. Rev. E **88**, 032109.

Ódor, G., and R. Pastor-Satorras, 2012, Phys. Rev. E **86**, 026117.

Olinky, R., and L. Stone, 2004, Phys. Rev. E **70**, 030902.

Oliveira, J. G., and A.-L. Barabasi, 2005, Nature (London) **437**, 1251.

Omic, J., A. Orda, and P. Van Mieghem, 2009, in *Proceedings of the IEEE INFOCOM 2009, 28th Annual Joint Conference of the IEEE Computer and Communications Societies* (IEEE, New York), pp. 1485–1493.

Onnela, J., J. Saramäki, J. Kertész, and K. Kaski, 2005, Phys. Rev. E **71**, 065103.

Onnela, J.-P., J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, 2007, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332.

Park, J., and M. E. J. Newman, 2003, Phys. Rev. E **68**, 026112.

Parshani, R., S. Carmi, and S. Havlin, 2010, Phys. Rev. Lett. **104**, 258701.

Pastor-Satorras, R., A. Vázquez, and A. Vespignani, 2001, Phys. Rev. Lett. **87**, 258701.
Pastor-Satorras, R., and A. Vespignani, 2001a, Phys. Rev. E **63**, 066117.
Pastor-Satorras, R., and A. Vespignani, 2001b, Phys. Rev. Lett. **86**, 3200.
Pastor-Satorras, R., and A. Vespignani, 2002a, Phys. Rev. E **65**, 035108.
Pastor-Satorras, R., and A. Vespignani, 2002b, Phys. Rev. E **65**, 036104.
Pastor-Satorras, R., and A. Vespignani, 2004, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England).
Payne, J. L., K. D. Harris, and P. S. Dodds, 2011, Phys. Rev. E **84**, 016110.
Peng, C., X. Jin, and M. Shi, 2010, Physica (Amsterdam) **389A**, 549.
Peng, S., and C. Li, 2009, in *International Conference on Computational Intelligence and Software Engineering, 2009. CiSE 2009* (IEEE, New York), pp. 1–4 [doi:10.1109/CISE.2009.5365379].
Perra, N., D. Balcan, B. Gonçalves, and A. Vespignani, 2011, PLoS One **6**, e23084.
Perra, N., A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, 2012, Phys. Rev. Lett. **109**, 238701.
Perra, N., B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, 2012, Sci. Rep. **2**, 469.
Pinto, P. C., P. Thiran, and M. Vetterli, 2012, Phys. Rev. Lett. **109**, 068702.
Poletto, C., S. Meloni, V. Colizza, Y. Moreno, and A. Vespignani, 2013, PLoS Comput. Biol. **9**, e1003169.
Poletto, C., M. Tizzoni, and V. Colizza, 2012, Sci. Rep. **2**, 476.
Prakash, B., D. Chakrabarti, N. Valler, M. Faloutsos, and C. Faloutsos, 2012, Knowledge and Information Systems **33**, 549.
Rattana, P., K. B. Blyuss, K. T. D. Eames, and I. Z. Kiss, 2013, Bull. Math. Biol. **75**, 466.
Ravasz, E., and A.-L. Barabási, 2003, Phys. Rev. E **67**, 026112.
Riley, S., 2007, Science **316**, 1298.
Risau-Gusman, S., and D. H. Zanette, 2009, J. Theor. Biol. **257**, 52.
Rocha, L. E. C., and V. D. Blondel, 2013, PLoS Comput. Biol. **9**, e1002974.
Rocha, L. E. C., F. Liljeros, and P. Holme, 2011, PLoS Comput. Biol. **7**, e1001109.
Rogers, E. M., 2010, *Diffusion of Innovations* (Simon and Schuster, New York).
Ross, S. M., 1996, *Stochastic Processes* (John Wiley & Sons, New York).
Rvachev, L., and I. Longini, 1985, Math. Biosci. **75**, 3.
Sahneh, F. D., C. Scoglio, and F. N. Chowdhury, 2013, in *American Control Conference (ACC), 2013* (AACC), pp. 2307–2312.
Sahneh, F. D., C. Scoglio, and P. Van Mieghem, 2013, IEEE/ACM Trans. Netw. **21**, 1609.
Salathé, M., and S. Bonhoeffer, 2008, J. R. Soc. Interface **5**, 1505.
Salathé, M., and J. H. Jones, 2010, PLoS Comput. Biol. **6**, e1000736.
Saldaña, J., 2008, Phys. Rev. E **78**, 012902.
Santos, A., J. M. F. Moura, and J. Xavier, 2013, arXiv:1306.6812.
Sattenspiel, L., and K. Dietz, 1995, Math. Biosci. **128**, 71.
Saumell-Mendiola, A., M. Á. Serrano, and M. Boguñá, 2012, Phys. Rev. E **86**, 026106.
Schneider, C. M., T. Mihaljev, S. Havlin, and H. J. Herrmann, 2011, Phys. Rev. E **84**, 061911.
Schumm, P., C. Scoglio, D. Gruenbacher, and T. Easton, 2007, in *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007* (Create-Net, Trento, Italy), pp. 201–208, 2nd ed.
Schwartz, N., R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, 2002, Phys. Rev. E **66**, 015104.
Schwartz, N., and L. Stone, 2013, Phys. Rev. E **87**, 042815.
Seidman, S. B., 1983, Soc. Networks **5**, 269.
Serrano, M. Á., and M. Boguñá, 2005, Phys. Rev. E **72**, 036133.
Serrano, M. A., and M. Boguñá, 2006, Phys. Rev. Lett. **97**, 088701.
Serrano, M. A., M. Boguñá, and R. Pastor-Satorras, 2006, Phys. Rev. E **74**, 055101.
Shalizi, C. R., and A. C. Thomas, 2011, Sociological Methods & Research **40**, 211.
Sharkey, K., 2008, J. Math. Biol. **57**, 311.
Sharkey, K., I. Kiss, R. Wilkinson, and P. Simon, 2015, Bull. Math. Biol. **77**, 614.
Sharkey, K. J., 2011, Theor. Popul. Biol. **79**, 115.
Shaw, L. B., 2008, Phys. Rev. E **77**, 066101.
Shaw, L. B., and I. B. Schwartz, 2010, Phys. Rev. E **81**, 046120.
Simon, P., M. Taylor, and I. Kiss, 2011, J. Math. Biol. **62**, 479.
Singh, P., S. Sreenivasan, B. K. Szymanski, and G. Korniss, 2013, Sci. Rep. **3**, 2330.
Small, M., and C. K. Tse, 2005, Int. J. Bifurcation Chaos Appl. Sci. Eng. **15**, 1745.
Solomonoff, R., and A. Rapoport, 1951, Bull. Math. Biophys. **13**, 107.
Son, S.-W., G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, 2012, Europhys. Lett. **97**, 16006.
Stanley, H. E., 1971, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford).
Starnini, M., A. Baronchelli, A. Barrat, and R. Pastor-Satorras, 2012, Phys. Rev. E **85**, 056115.
Starnini, M., A. Machens, C. Cattuto, A. Barrat, and R. Pastor-Satorras, 2013, J. Theor. Biol. **337**, 89.
Starnini, M., and R. Pastor-Satorras, 2013, Phys. Rev. E **87**, 062807.
Starnini, M., and R. Pastor-Satorras, 2014, Phys. Rev. E **89**, 032807.
Stauffer, A. O., and V. C. Barbosa, 2006, Phys. Rev. E **74**, 056105.
Stauffer, D., and A. Aharony, 1994, *Introduction to Percolation Theory* (Taylor & Francis, London), 2nd ed.
Stauffer, D., and M. Sahimi, 2005, Phys. Rev. E **72**, 046128.
Stehle, J., *et al.*, 2011, BMC Med. **9**, 87.
Sudbury, A., 1985, J. Appl. Probab. **22**, 443.
Takaguchi, T., N. Masuda, and P. Holme, 2013, PLoS One **8**, e68629.
Tang, J., S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, 2010, Phys. Rev. E **81**, 055101.
Tanimoto, S., 2011, arXiv:1103.1680.
Tijms, H., 2003, *A First Course in Stochastic Models* (Wiley, Chichester).
Tizzoni, M., P. Bajardi, C. Poletto, J. Ramasco, D. Balcan, B. Goncalves, N. Perra, V. Colizza, and A. Vespignani, 2012, BMC Med. **10**, 165.
Trapman, P., 2007, Theor. Popul. Biol. **71**, 160.
Trpevski, D., W. K. S. Tang, and L. Kocarev, 2010, Phys. Rev. E **81**, 056102.
Tunc, I., M. S. Shkarayev, and L. B. Shaw, 2013, J. Stat. Phys. **151**, 355.
Valdez, L., P. A. Macri, and L. A. Braunstein, 2012a, Phys. Rev. E **85**, 036108.
Valdez, L. D., P. A. Macri, and L. A. Braunstein, 2012b, PLoS One **7**, e44188.
van de Bovenkamp, R., F. Kuipers, and P. Van Mieghem, 2014, Phys. Rev. E **89**, 042818.
van Kampen, N. G., 1981, *Stochastic Processes in Chemistry and Physics* (North-Holland, Amsterdam).
Van Mieghem, P., 2011, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, England).

Van Mieghem, P., 2012a, Europhys. Lett. **97**, 48004.
Van Mieghem, P., 2012b, Comput. Commun. **35**, 1494.
Van Mieghem, P., 2013, arXiv:1310.3980.
Van Mieghem, P., 2014a, arXiv:1402.1731.
Van Mieghem, P., 2014b, *Performance Analysis of Complex Networks and Systems* (Cambridge University Press, Cambridge, England).
Van Mieghem, P., N. Blenn, and C. Doerr, 2011, Eur. Phys. J. B **83**, 251.
Van Mieghem, P., and E. Cator, 2012, Phys. Rev. E **86**, 016116.
Van Mieghem, P., and J. Omic, 2008, arXiv:1306.2588.
Van Mieghem, P., J. Omic, and R. E. Kooij, 2009, IEEE/ACM Trans. Netw. **17**, 1.
Van Mieghem, P., F. D. Sahneh, and C. Scoglio, 2014, in *Proceedings of the 53rd IEEE Conference on Decision and Control, CDC'14* (IEEE, New York).
Van Mieghem, P., and R. van de Bovenkamp, 2013, Phys. Rev. Lett. **110**, 108701.
Van Mieghem, P., H. Wang, X. Ge, S. Tang, and F. A. Kuipers, 2010, Eur. Phys. J. B **76**, 643.
Van Segbroeck, S., F. C. Santos, and J. M. Pacheco, 2010, PLoS Comput. Biol. **6**, e1000895.
Vazquez, A., 2007, J. Theor. Biol. **245**, 125.
Vazquez, A., B. Rácz, A. Lukács, and A.-L. Barabási, 2007, Phys. Rev. Lett. **98**, 158702.
Vázquez, A., M. Boguñá, Y. Moreno, R. Pastor-Satorras, and A. Vespignani, 2003, Phys. Rev. E **67**, 046111.
Vázquez, A., and Y. Moreno, 2003, Phys. Rev. E **67**, 015101.
Vázquez, A., R. Pastor-Satorras, and A. Vespignani, 2002, Phys. Rev. E **65**, 066130.
Ver Steeg, G., R. Ghosh, and K. Lerman, 2011, in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (AAAI Press, Menlo Park, CA), pp. 377–384.
Vespignani, A., 2009, Science **325**, 425.
Vespignani, A., 2012, Nat. Phys. **8**, 32.
Viboud, C., O. Bjornstad, D. Smith, L. Simonsen, M. Miller, and B. Grenfell, 2006, Science **312**, 447.
Vojta, T., 2006, J. Phys. A **39**, R143.
Volz, E., 2008, J. Math. Biol. **56**, 293.
Volz, E., and L. A. Meyers, 2009, J. R. Soc. Interface **6**, 233.
Wang, Y., D. Chakrabarti, C. Wang, and C. Faloutsos, 2003, in *22nd International Symposium on Reliable Distributed Systems (SRDS'03)* (IEEE Computer Society, Los Alamitos, CA), pp. 25–34.
Warren, C. P., L. M. Sander, and I. M. Sokolov, 2002, Phys. Rev. E **66**, 056105.
Wasserman, S., and K. Faust, 1994, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, England).
Watts, D. J., 2002, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766.
Watts, D. J., R. Muhamad, D. C. Medina, and P. S. Dodds, 2005, Proc. Natl. Acad. Sci. U.S.A. **102**, 11157.
Watts, D. J., and S. H. Strogatz, 1998, Nature (London) **393**, 440.
Webb, J. N., 2007, *Game Theory: Decisions, Interaction and Evolution*, Springer Undergraduate Mathematics Series (Springer-Verlag, London).
Weber, S., and M. Porto, 2007, Phys. Rev. E **76**, 046111.
Wen, L., and J. Zhong, 2012, Nonlinear Anal.: Real World Appl. **13**, 967.
Weng, L., J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini, 2013, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13* (ACM, New York), pp. 356–364.
Wilf, H. S., 2006, *Generating Functionology* (A. K. Peters, Ltd., Natick, MA).
Wilkinson, R. R., and K. J. Sharkey, 2014, Phys. Rev. E **89**, 022808.
Wu, X., and Z. Liu, 2008, Physica (Amsterdam) **387A**, 623.
Yagan, O., D. Qian, J. Zhang, and D. Cochran, 2013, IEEE Journal on Selected Areas in Communications **31**, 1038.
Yan, G., Z.-Q. Fu, J. Ren, and W.-X. Wang, 2007, Phys. Rev. E **75**, 016108.
Yang, Z., and T. Zhou, 2012, Phys. Rev. E **85**, 056106.
Yeomans, J. M., 1992, *Statistical Mechanics of Phase Transitions* (Oxford University Press, Oxford).
Youssef, M., R. E. Kooij, and C. Scoglio, 2011, J. Comput. Sci. **2**, 286.
Youssef, M., and C. Scoglio, 2011, J. Theor. Biol. **283**, 136.
Zanette, D., 2001, Phys. Rev. E **64**, 050901.
Zanette, D. H., and S. Risau-Gusmán, 2008, J. Biol. Phys. **34**, 135.
Zeng, A., and C.-J. Zhang, 2013, Phys. Lett. A **377**, 1031.
Zhao, H., and Z. Gao, 2007, Europhys. Lett. **79**, 38002.
Zhou, J., Z. Liu, and B. Li, 2007, Phys. Lett. A **368**, 458.