Cumulative Merging Percolation and the Epidemic Transition of the Susceptible-Infected-Susceptible Model in Networks

Claudio Castellano[1] and Romualdo Pastor-Satorras[2]

[1]*Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy*
[2]*Departament de Física, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*

We consider cumulative merging percolation (CMP), a long-range percolation process describing the iterative merging of clusters in networks, depending on their mass and mutual distance. For a specific class of CMP processes, which represents a generalization of degree-ordered percolation, we derive a scaling solution on uncorrelated complex networks, unveiling the existence of diverse mechanisms leading to the formation of a percolating cluster. The scaling solution accurately reproduces universal properties of the transition. This finding is used to infer the critical properties of the susceptible-infected-susceptible model for epidemics in infinite and finite power-law distributed networks. Here, discrepancies between analytical approaches and numerical results regarding the finite-size scaling of the epidemic threshold are a crucial open issue in the literature. We find that the scaling exponent assumes a nontrivial value during a long preasymptotic regime. We calculate this value, finding good agreement with numerical evidence. We also show that the crossover to the true asymptotic regime occurs for sizes much beyond currently feasible simulations. Our findings allow us to rationalize and reconcile all previously published results (both analytical and numerical), thus ending a long-standing debate.

I. INTRODUCTION

Percolation and epidemic spreading are among the most interesting processes unfolding on complex network substrates, and their investigation has attracted a huge interest in the past 20 years [1–5]. One of the most successful achievements of this endeavor is the realization that the properties of one of the fundamental models for epidemics without a steady state, the susceptible-infected-recovered (SIR) dynamics [6], can be mapped onto bond percolation [7–9]. This connection has permitted the application to the SIR model of the powerful tools devised for percolation, leading to a full understanding of this epidemic process [8,10–13]. For the other fundamental class of epidemic dynamics, allowing for a steady, endemic state, whose simplest representative is the susceptible-infected-susceptible (SIS) model [6], no direct mapping to a percolative framework is available, and theoretical progress has been slower. In the SIS model, susceptible individuals acquire the disease at rate β through any edge connected to an infected individual, while infected individuals spontaneously heal with rate μ. The epidemic threshold λ_c defines the value of the ratio $\lambda = \beta/\mu$ separating a healthy (absorbing) phase from an endemic one with everlasting activity. Initial work showed that degree heterogeneity leads to disruptive effects on scale-free networks [14], namely, a vanishing threshold in networks with power-law degree distribution $P(k) \sim k^{-\gamma}$ and $\gamma \leq 3$ [15,16]. Later efforts have shifted toward less heterogeneous networks, those with $\gamma > 3$ [3].

The quenched mean-field (QMF) theory [17–19] predicts a vanishing threshold $\lambda_c \to 0$ in the infinite network-size limit for any value of γ [20], due to the existence of hubs able to sustain the epidemic for long times only by interacting with their direct neighbors [21]. It was later pointed out that, at the QMF level, the localization of activity around these hubs implies the existence, for small values of λ, of long-lived, but not stationary, states [22,23]. Important progress in this debate is provided in Ref. [24], where it is shown that a genuine non-mean-field effect, mutual reinfection among distant hubs, is the key mechanism triggering the appearance of an endemic stationary state for any λ in networks with $\gamma > 5/2$. Numerical evidence corroborates this picture, showing that the position of the effective threshold tends to zero with network size for any γ. However, the decay observed is slower than

the one predicted by the QMF theory [24–26] and, moreover, in contradiction with recent mathematical results derived by Huang and Durrett [27] and Mountford, Valesin, and Yao [28]. An additional puzzling question in this area is the striking disagreement between the exact mathematical prediction for the singular behavior of the prevalence ($\rho \sim \lambda^{2\gamma-3}$, apart from logarithmic corrections) [28] and numerical simulations exhibiting a much faster growth. This lack of a precise agreement between analytics and numerics represents one standing issue in our understanding of epidemic processes on complex topologies.

A precise mathematical formulation of the mutual reinfection process was recently proposed by Ménard and Singh [29]. They introduce the cumulative merging percolation (CMP) process, a long-range site percolation process [30] aimed at describing the geometry of the sets where SIS epidemics survive for a long time on a network. The presence of a CMP giant component corresponds to the existence of an endemic SIS stationary state, so that the calculation of the CMP threshold allows one to locate also the position of the SIS epidemic transition [31].

In this paper, we contribute to the current state-of-the-art in this area in two ways. First, we consider a generalized version of the CMP process proposed in Ref. [29], and we present a scaling theory for its nontrivial behavior. This theory—which provides a clear understanding of competing physical mechanisms, critical properties, crossover scales, and finite-size effects—is general and can be related with other processes. Second, we apply the results of the first part to SIS dynamics, obtaining in this way for the first time a full understanding of the critical properties of the model. In particular, our theory predicts that the asymptotic behavior in the limit of very large networks (derived exactly in Refs. [27,28] and reassuringly recovered by our approach) can be observed only for huge system sizes, out of reach for present computer resources. We show instead that, for network sizes that can be currently simulated, a preasymptotic regime holds, whose nontrivial properties are determined, providing a prediction for the finite-size scaling of the SIS epidemic threshold in agreement with (previously unexplained) numerical results. Our work reconciles in a comprehensive way the different theories proposed to interpret the behavior of the SIS model, placing them in the proper context regarding the network size considered, and thus ends a long debate between the physics and mathematics communities.

The paper is organized as follows: In Sec. II, we define the cumulative merging percolation process which is the subject of our study. Section III presents a scaling solution of this model, whose behavior in finite networks is discussed in Sec. IV. A numerical check of the scaling solution is provided in Sec. V. In Sec. VI, we apply the results obtained to the SIS epidemic model, backing up our conclusions by comparison with existing numerical simulations. Finally, in Sec. VII, we summarize our main results

and discuss the interesting perspectives they open. Several Appendixes provide some detailed analytical calculations and additional information.

II. CUMULATIVE MERGING PERCOLATION PROCESS

We consider a generalization of the cumulative merging process proposed in Ref. [29], defined along the following lines. In a given network, composed by N nodes, each node i is *active* with probability p_i. Inactive nodes do not play any role apart from determining the topological distances between pairs of active nodes (see below). Each active node i defines a cluster of size 1, associated with an initial mass $m_i^{(0)}$. Starting with these initial clusters, an iterative process takes place whose elementary step is the merging of a pair of clusters into a single one. Two clusters, α and β, are merged in a single cluster if there are at least a node i_α in α and a node j_β in β, such that

$$d_{i_\alpha, j_\beta} \leq \min\{r(m_\alpha), r(m_\beta)\}, \qquad (1)$$

where $d_{i,j}$ is the topological distance between nodes i and j and $r(m) \geq 1$ is an interaction range associated to a cluster of mass m. The mass of the merged cluster is the sum of the masses of the original clusters: $m_{\alpha+\beta} = m_\alpha + m_\beta$. The iteration of this procedure converges to a limiting partition of the network that does not depend on the order in which the merging is performed [32]. Notice that if $p_i = p$ and $r(m) = 1$, CMP coincides with random site percolation [1]. It is important to remark that Eq. (1) implies that two clusters merge only if each one of them is within the interaction range of the other: An asymmetric situation,
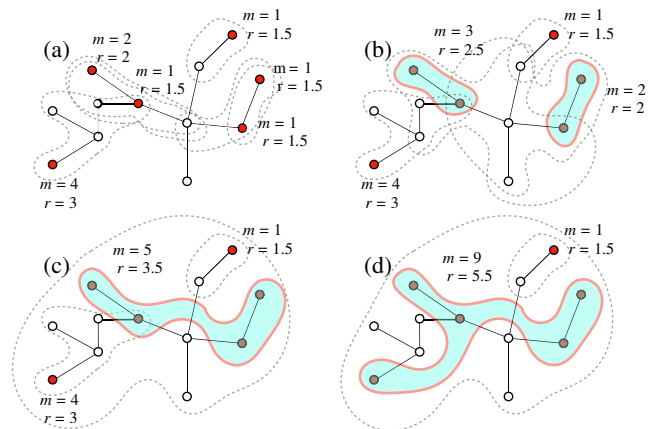

FIG. 1. Schematic illustration of the CMP process for $r(m) = 1 + m/2$. Filled nodes are active; empty nodes are inactive. Areas bordered by dashed lines are interaction domains of active nodes or clusters. Clusters are indicated by solid lines surrounding a filled area. Notice in (c) and (d) the isolated active node on the upper right corner, which is within the interaction range of the large cluster but cannot be merged as it has $r < 2$.

with a massive cluster interacting with a far and small cluster, does not lead to merging. In Fig. 1, we present a graphical illustration of the mechanism of the CMP process. We stress again that a cluster is defined as a set of (only) active nodes resulting from the iteration of merging events. Nodes in the same cluster must belong to the same connected component of the underlying network, but they do *not* need to form a connected component by themselves, as is clear from Figs. 1(c) and 1(d).

The connection between CMP and the mutual reinfection of distant hubs in the SIS epidemics is operated by taking as active nodes the hubs able to independently sustain the epidemic [29]; see Appendix A for a detailed description.

III. SCALING THEORY FOR CUMULATIVE MERGING PERCOLATION

Let us focus now on a specific yet broad class of CMP processes, where nodes are active if their degree is larger than a threshold value k_a: $p_i \equiv p(k_i) = \Theta(k_i - k_a)$. In an uncorrelated network with degree distribution $P(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$ in the continuous approximation, where $k_{\min}$ is the minimum degree, the fraction of active nodes is

$$\frac{N_a}{N} = \int_{k_a}^{\infty} dk P(k) = \left(\frac{k_a}{k_{\min}}\right)^{1-\gamma}. \qquad (2)$$

We are interested in understanding the possible existence of a CMP giant component as a function of k_a, in particular, in the limit $k_a \to \infty$, when only a small fraction of nodes is active.

A. The case $r(m) = 1$: Degree-ordered percolation

Let us consider first the case $r(m) = 1$, i.e., only nearest neighbors can form clusters. In this case, the CMP process defined above coincides with the degree-ordered percolation (DOP) process proposed in Ref. [23] (coinciding with the limit $\alpha \to -\infty$ in Ref. [33]). For a node of degree k, the probability that a given neighbor is active is

$$P_a(k) = \int_{k_a}^{\infty} dk' P(k'|k), \qquad (3)$$

where $P(k'|k)$ is the conditional probability that a neighbor of a node k has degree k' [34]. For uncorrelated networks, $P(k'|k) = (k'P(k')/\langle k \rangle)$ [34]; thus, we have $P_a = (k_a/k_{\min})^{2-\gamma}$, independent of k. The mean number of active neighbors of a node of degree k is kP_a; therefore, the inverse of P_a,

$$k_c = \left(\frac{k_{\min}}{k_a}\right)^{2-\gamma}, \qquad (4)$$

defines a degree scale separating nodes likely to have many active neighbors $k/k_c \gg 1$ from those likely to be *isolated*,

i.e., not in direct contact with any active node. The average number of active neighbors for each active node is

$$\frac{N}{N_a} \int_{k_a}^{\infty} dk P(k) k P_a = \frac{\gamma - 1}{\gamma - 2} k_a P_a \sim k_a^{3-\gamma}. \qquad (5)$$

For $\gamma < 3$, this quantity diverges as k_a grows: Each active node has a very large number of active neighbors, so that all of them belong to a connected giant component for any k_a [23,33], and the relative size S of the giant component is simply given by the fraction of active nodes

$$S_{\text{DOP}} = \frac{N_a}{N} = \left(\frac{k_a}{k_{\min}}\right)^{1-\gamma}. \qquad (6)$$

For $\gamma > 3$, instead, the average number of active neighbors of an active node decreases with k_a and tends to zero in the limit $k_a \to \infty$. This result indicates that a degree-ordered percolation giant component (DOPGC) can exist only up to a finite threshold value, in agreement with Refs. [23,33]. It is useful to discuss the behavior of the order parameter S_{DOP} as a function of k_a in this case. For $k_a = k_{\min}$, $k_c = 1$. Hence, even for $\gamma > 3$, there is an interval of k_a values such that $k_a/k_c > 1$. This regime occurs up to a value $k_a = k_0^*$ determined by the condition $k_c(k_0^*) = k_0^*$, yielding

$$k_0^* = k_{\min}^{(\gamma-2)/(\gamma-3)}. \qquad (7)$$

Notice that, for $\gamma = 3.2$ and $k_{\min} = 3$, $k_0^* = 729$, a quite large value, while it decays quickly for increasing γ: For $\gamma = 3.5$, it is already $k_0^* = 27$. In this regime, the situation is similar to the case $\gamma < 3$, with practically all active nodes belonging to the DOPGC and $S_{\text{DOP}} \approx N_a/N \sim k_a^{1-\gamma}$. However, one must notice that, even if $k_a/k_c > 1$, this ratio is not very large, as its maximum value is $k_{\min}$, corresponding to $k_a = k_{\min}$. Therefore, one never observes the scaling predicted by Eq. (6); as soon as k_a is increased, one immediately starts to see the transition to a different regime, where $k_a/k_c < 1$. In this second regime, a giant component still exists, but some active nodes are isolated (not directly connected to other active nodes) and others are nonisolated but form small clusters. The set of all active nodes is therefore composed by three classes:

(1) nonisolated nodes belonging to the DOPGC;
(2) nonisolated nodes belonging to small clusters; and
(3) isolated nodes, which necessarily do not belong to the DOPGC.

As k_a increases, a growing fraction of active nodes passes from the first category to the other two, and the order parameter $S_{\text{DOP}} = N_{\text{GC}}/N$ decreases faster than the fraction of active nodes N_a/N (see Fig. 2). At the threshold, the fraction of nonisolated nodes belonging to the DOPGC vanishes.
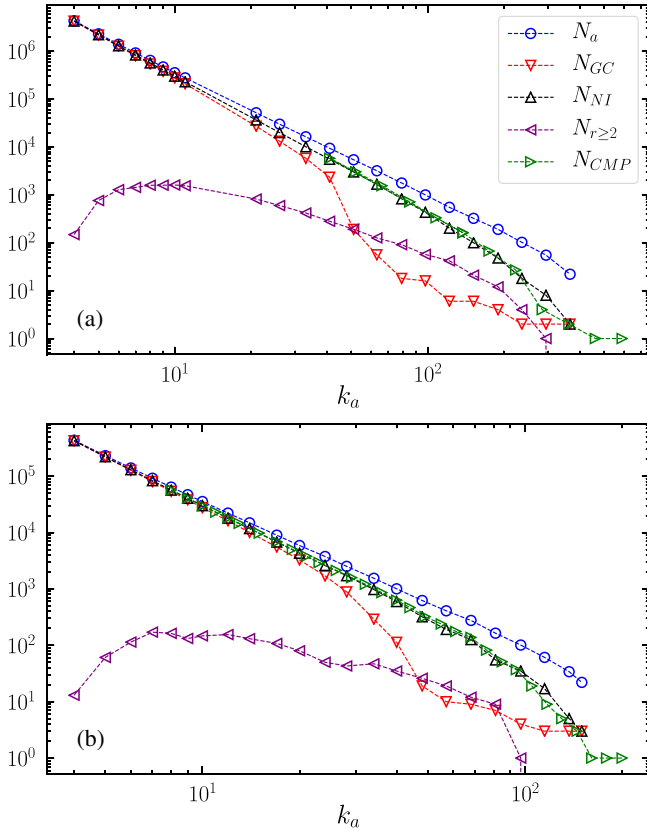
FIG. 2. Dependence on k_a of the number N_a of active nodes, the number N_{GC} of nodes in the DOP giant component, the number N_{NI} of nonisolated nodes, the number $N_{r \geq 2}$ of isolated nodes with range $r \geq 2$, and the number N_{CMP} of nodes in the CMP giant component. Results are for power-law networks with $\gamma = 3.5$, $k_{min} = 3$, and size $N = 10^7$ (a) and $N = 10^6$ (b), generated using the uncorrelated configuration model (UCM) [35].

The calculation of the behavior of S_{DOP} in this regime and of the transition point is a nontrivial task. It is important to observe that N_{NI}, the number of nonisolated nodes, which upper bounds the number N_{GC} of nodes belonging to the giant component, keeps decaying with the same exponent even well above the DOP transition (see Fig. 2).

B. The case of growing $r(m)$ for $\gamma < 3$

Let us turn now to the more generic case where $r(m)$ grows as a function of m. The range of interaction grows with its mass, so that, if $r(m) \geq 2$, clusters of nodes can merge even if not in direct contact. In this case, it is clear that, for a given value of k_a, the giant component of the DOP process is a subset of the giant component of the full CMP process (CMPGC). Thus, for $\gamma < 3$, the CMPGC is again given by the whole set of active nodes and has, therefore, a relative size

$$S = \frac{N_a}{N} = \left(\frac{k_a}{k_{min}} \right)^{1-\gamma}. \tag{8}$$

C. The case of growing $r(m)$ for $\gamma > 3$

In this case, for large k_a, the DOPGC vanishes asymptotically and nonisolated active nodes form DOP clusters of a small size. Still, an extensive CMPGC could be induced by long-range merging of clusters or nodes which cannot be joined in a DOP process, as they are separated by distances larger than 1. Whether these long-range mergings take place or not depends, of course, on the particular choice of the mass m and of the form of the interaction range. Inspired by Ref. [29], here we focus on the case of initial masses equal to node degrees $m_i^{(0)} = k_i$ (so that the total mass of a cluster is the sum of the degrees of the active nodes forming it) and of an interaction range of the form $r(m) = m/k_a$. This case is a particular case of a generic CMP process with $r(m) = f(m/k_a)$, where $f(z) = z^\alpha$, with $\alpha > 0$, so that active nodes with the smallest degree have a range exactly equal to 1. We defer to a future work a comprehensive analysis of this model for $\alpha \neq 1$.

In the present setting, we identify two competing mechanisms leading to the formation of a CMP giant component. The first is an extension of DOP percolation, based on the merging of DOP clusters separated by distances larger than 1. The second involves the buildup of CMP clusters formed by isolated nodes interacting at a large distance. We now discuss the two mechanisms in detail.

1. First mechanism: Extended DOP mechanism

For very small k_a close to k_{min}, CMP is clearly equivalent to the first regime for DOP with $S \approx N_a/N$. Upon increasing k_a, above the crossover scale k_0^*, DOP enters the second regime with an increasing presence of isolated nodes and nodes belonging to small DOP clusters. CMP and DOP behaviors start to diverge at this point, because some nodes, even if they are not directly connected to the DOPGC, are at distance 2 from it and, thus, can join the CMPGC if their interaction range is at least 2. In particular, this situation occurs for all small DOP clusters: As their aggregate degree is $k_{agg} \geq 2k_a$, they necessarily have a range of interaction $r \geq 2$. For this reason, in this regime all N_{NI} nonisolated nodes belong to the CMPGC. This result is clearly verified in Fig. 2. Notice that N_{NI}/N is finite even well beyond the DOP threshold. In this limit, the formation of the CMPGC is still triggered by the largest DOP cluster (that does not percolate). For any value of γ, there are always nodes in the network with $k > k_c \gg k_a$. They form local clusters with a large interaction range that progressively incorporate other small clusters giving rise to a CMPGC, even if no DOPGC is present. To calculate N_{NI}, we consider the probability that an active node of degree k has at least one neighboring active node:

$$P_{NI}(k) = 1 - (1 - P_a)^k \approx 1 - e^{-k/k_c}. \tag{9}$$

The total fraction $N_{\rm NI}/N$ of nonisolated active nodes in a power-law distributed network is then

$$\frac{N_{\rm NI}}{N} = \int_{k_a}^{\infty} dk' P(k') P_{\rm NI}(k')$$

$$= (\gamma - 1)k_{\min}^{\gamma-1} \left[\frac{k_a^{1-\gamma}}{\gamma - 1} - k_c^{1-\gamma} \Gamma\left(1 - \gamma, \frac{k_a}{k_c}\right)\right], \quad (10)$$

where $\Gamma(a, z)$ is the incomplete Gamma function [36].

In this second regime, not only small DOP clusters, but also isolated active nodes can join the CMPGC, provided they have degree $k \geq 2k_a$ so that their range is $r \geq 2$. We denote their number as $N_{r\geq 2}$. The total fraction of isolated nodes with range $r \geq 2$ is

$$\frac{N_{r\geq 2}}{N} = \int_{2k_a}^{\infty} dk' P(k')[1 - P_{\rm NI}(k')] \quad (11)$$

$$= (\gamma - 1)k_{\min}^{\gamma-1} k_c^{1-\gamma} \Gamma\left(1 - \gamma, \frac{2k_a}{k_c}\right). \quad (12)$$

Overall, the CMP order parameter in this regime is, therefore,

$$S_1 \approx \frac{N_{\rm NI}}{N} + \frac{N_{r\geq 2}}{N}. \quad (13)$$

For $k_a \to k_{\min}$, one has $k_a > k_c$, and the first contribution in Eq. (13) is larger than the second, for any γ. For large k_a, instead, one can expand the Γ functions for small k_a/k_c, finding

$$\frac{N_{\rm NI}}{N} = \left(\frac{k_a}{k_{\min}}\right)^{1-\gamma} \left[\frac{\gamma - 1}{\gamma - 2} \frac{k_a}{k_c}\right] \sim k_a^{2(2-\gamma)} \quad (14)$$

and

$$\frac{N_{r\geq 2}}{N} = \left(\frac{2k_a}{k_{\min}}\right)^{1-\gamma} \left[1 - \frac{\gamma - 1}{\gamma - 2} \frac{2k_a}{k_c}\right] \sim k_a^{1-\gamma}. \quad (15)$$

The exponent of $N_{\rm NI}$ is, in absolute value, larger than the one of $N_{r\geq 2}$; hence, the first contribution dominates up to a crossover scale

$$k_1^* = \left[\frac{(\gamma - 2)}{(\gamma - 1)} \frac{2^{(1-\gamma)}}{(1 + 2^{2-\gamma})} k_{\min}^{(2-\gamma)}\right]^{1/(3-\gamma)}. \quad (16)$$

The conclusion of this line of reasoning is that for $k_a \ll k_1^*$ the size of the CMPGC decays as

$$S_1 \approx \frac{N_{\rm NI}}{N} \sim k_a^{2(2-\gamma)} \quad (17)$$

followed by a crossover to $S_1 \approx (N_{r\geq 2}/N) \sim k_a^{1-\gamma}$. The crossover scale k_1^* decreases rapidly with γ, but, since the
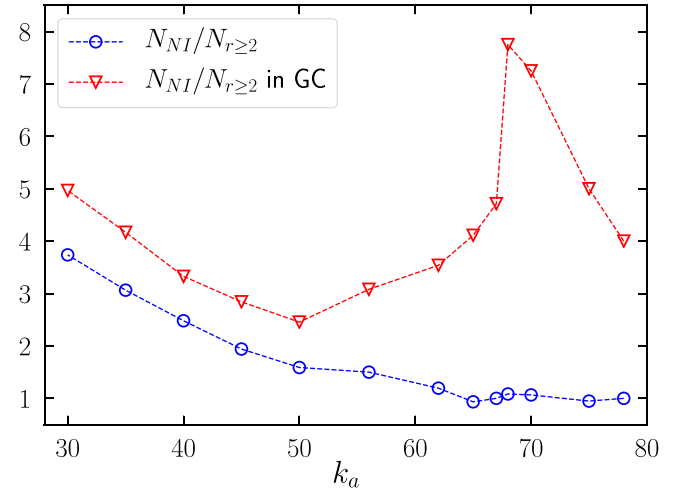

FIG. 3. Ratio $N_{\rm NI}/N_{r\geq 2}$ between the two terms in Eq. (13) evaluated on UCM networks with $\gamma = 4$, $k_{\min} = 3$, and size $N = 10^7$, computed over all nodes, and restricted to nodes belonging to the CMPGC.

maximum degree in a network grows as $N^{1/(\gamma-1)}$, the minimum network size necessary to have a sufficiently large maximum degree $k_{\max} = k_1^*$ is always larger than $N \approx 4.2 \times 10^5$ (the minimum occurring for $\gamma \approx 5$ for $k_{\min} = 3$). Hence, it should be possible to observe the crossover on large networks (although $k_{\max}$ grows very slowly with N; hence, one needs networks of a size much larger than 10^5 nodes to have a still limited range of k_a values). As a matter of fact, we do not observe such a crossover.

This result happens because, as k_a grows, the extended DOP mechanism becomes less and less effective. DOP clusters become smaller and smaller, and the distances among them (and between isolated active nodes and them) increase: It is no longer sufficient to have $r = 2$ to join the CMP giant component. For even larger k_a, it is not even sufficient to have $r = 3$ or $r = 4$ and so on. This effect suppresses both terms in Eq. (13), but the second term is most affected, as can be seen in Fig. 3, where we compare the ratio of the first and second terms in Eq. (13) (which becomes 1 at the crossover scale k_1^*) and the same ratio restricted to nodes belonging to the CMPGC. We observe that the latter is always larger than the former and does not seem to go to 1 for large k_a. This result implies that, in practice, S_1 behaves as predicted by Eq. (17) even for values of k_a larger than the crossover scale k_1^* estimated in Eq. (16).

A more important consequence of the asymptotic ineffectiveness of the extended DOP mechanism is that it cannot work for arbitrarily large k_a. A different mechanism governs the formation of the CMPGC in the limit $k_a \to \infty$.

2. Second mechanism: Merging of distant isolated nodes

Nodes with degree $k_a \leq k \ll k_c$ have on average a very small number of active nearest neighbors, as

$kP_a = k/k_c \ll 1$. Hence, they are typically isolated. However, if k is large enough, they may still have a large interaction range and may merge with other distant nodes. To analyze this process in detail, let us denote as $d(k)$ the mean distance between a node of degree k and the closest node of degree at least k. In the limit of a large network size, this distance is (see Appendix B for an analytical derivation)

$$d(k) \approx 1 + \frac{\gamma - 3}{\ln(\kappa)} \ln\left(\frac{k}{k_{\min}}\right), \qquad (18)$$

where $\kappa = \langle k^2 \rangle / \langle k \rangle - 1$ is the network branching factor. Since the interaction range of a node grows linearly with its degree k, it grows faster than the distance to its closest peer. Hence, there exists a degree k_x such that

$$r(k_x) = d(k_x) \qquad (19)$$

and, for any $k > k_x$, $r(k) > d(k)$. As a consequence, nodes with $k > k_x$ have an interaction range larger (on average) than their mutual topological distance. They can thus merge in pairs with an even larger interaction range, and the process repeats itself, leading to the formation of a CMPGC, comprising all nodes with a degree larger than k_x. If we write k_x in the form $k_x = \omega k_a$, the condition (19) implies $\omega = d(\omega k_a)$, which, inserting the explicit expression of $d(k)$, becomes

$$\omega \approx 1 + \frac{\gamma - 3}{\ln(\kappa)} \ln\left(\frac{\omega k_a}{k_{\min}}\right). \qquad (20)$$

Neglecting constants and terms of the order of $\ln[\ln(k_a)]$, the size of the giant component according to this mechanism scales then as

$$S_2 \approx \left(\frac{k_x}{k_{\min}}\right)^{1-\gamma} = \omega^{1-\gamma}\left(\frac{k_a}{k_{\min}}\right)^{1-\gamma} \qquad (21)$$

$$= \left[\frac{\gamma - 3}{\ln(\kappa)} \ln\left(\frac{k_a}{k_{\min}}\right)\right]^{(1-\gamma)}\left(\frac{k_a}{k_{\min}}\right)^{1-\gamma}, \qquad (22)$$

showing, thus, a power-law decay times a logarithmic correction.

In absolute value, the leading exponent in S_2 is smaller than the exponent in S_1. Therefore, we expect this second mechanism to dominate asymptotically, but after a crossover preceded by a scaling regime where the size of the CMPGC is given by Eq. (17). The position k_2^* of the crossover is estimated by numerically solving the equation $S_1(k_2^*) = S_2(k_2^*)$. Figure 4 shows how this quantity decreases with the exponent γ. However, in order to observe such a crossover, one must consider networks much larger than $N_2^* = k_2^{*(\gamma-1)}$. These values are huge for any γ (much larger than 10^9 nodes in the best case), leading
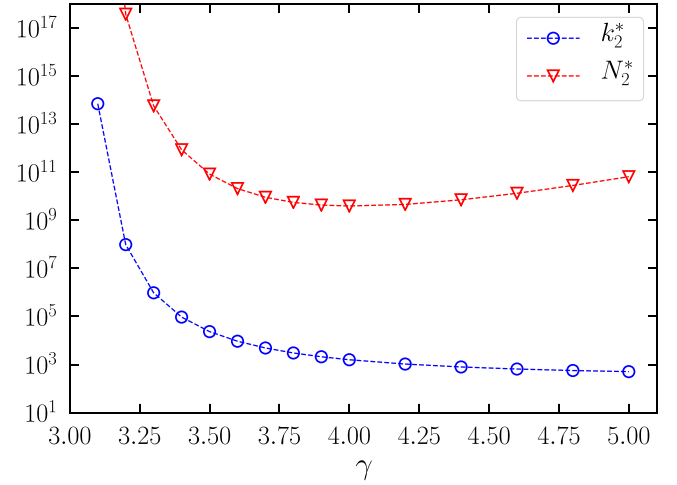

FIG. 4. Values of the crossover degree k_2^* and the minimal network size $N_2^* = k_2^{*(\gamma-1)}$ as a function of γ. In order to observe the crossover, networks of size $N \gg N_2^*$ should be considered.

to the conclusion that only the first regime can be observed in currently feasible simulations.

The present analysis can be extended also to the case of networks with a stretched exponential degree distribution, predicting an asymptotic stretched exponential dependence of S on k_a. See Appendix C for details.

IV. FINITE-SIZE EFFECTS

So far, we consider infinitely large networks, thus assuming that all degree classes, up to infinity, exist. When the network size is finite, only degrees up to the maximum value $k_{\max}(N)$, growing as $N^{1/(\gamma-1)}$, are present [37]. The CMP behavior for the infinite network (i.e., there is a CMPGC for any k_a) holds as long as $k_{\max}$ is larger than the degree scale involved in the formation of the CMPGC.

For $\gamma < 3$, it is sufficient to have active nodes for observing an extensive CMPGC. Hence, the only finite-size effect trivially appears for $k_a > k_{\max}(N)$: In such a case, there are no more active nodes in the system and $S \approx 0$. The finite-size effective threshold is $k_a^c = k_{\max}(N)$.

On the contrary, for $\gamma > 3$, finite-size effects are less trivial. The presence of active nodes is not sufficient to give rise to a CMPGC. One needs the presence of nodes with $k > k_c$ (first mechanism) or $k > k_x$ (second mechanism). Notice that, since k_c grows as a power of k_a with an exponent larger than 1, while k_x grows logarithmically, asymptotically $k_c \gg k_x$. Different scalings of the finite-size effective threshold are possible, depending on whether the maximum degree $k_{\max}(N)$ is larger or smaller than the crossover degree k_2^*.

If $k_{\max}(N) > k_2^*$, finite-size effects appear during the regime where the formation of the CMPGC is governed by the second mechanism. In this case, the asymptotic behavior $S \approx S_2$ ends (i.e., $S \approx 0$) when the relevant degree

scale k_x (growing with k_a) becomes larger than $k_{max}(N)$. In such a case, there are active nodes in the system, but neither of the two mechanisms for the formation of the giant component is at work. The effective threshold in this case is given by the condition $k_x = k_{max}(N)$, implying asymptotically

$$k_a^c \sim \frac{\ln(\kappa)}{\gamma - 3} \frac{k_{max}(N)}{\ln[k_{max}(N)]}. \tag{23}$$

If, instead, $k_{max}(N) < k_2^*$, finite-size effects start to appear already during the preasymptotic regime where the first mechanism rules. As soon as $k_c > k_{max}(N)$, the behavior $S \approx S_1$ ends. The effective threshold is thus given by the condition $k_c = k_{max}(N)$, implying

$$k_a^c = k_{min} k_{max}^{1/(\gamma-2)}. \tag{24}$$

Notice that after this effective threshold the order parameter does not go to $S \approx 0$, as there is still an interval of k_a values such that $k_x < k_{max}(N) < k_c$. In this regime, the first mechanism is no longer operative; still, the second is at work, but since $N < N_2^*$, it cannot lead to a macroscopic giant component.

V. NUMERICAL TEST

We test the correctness of the scaling analysis performed in the previous sections by means of numerical simulations of the CMP process with $r(m) = m/k_a$. In Figs. 5(a) and 5(b), we report, as a function of k_a, the fraction S of nodes in the largest CMP cluster for $\gamma < 3$ on uncorrelated configuration model networks (UCM) [35] of various sizes. The plot shows the presence of a CMPGC, including a fraction of active nodes independent of the system size N. The scaling of S with k_a is in excellent agreement with the prediction of Eq. (8). Finite-size effects are also apparent and perfectly agree with the prediction formulated above: The effective threshold occurs for $k_a = k_{max}(N) = N^{1/2}$. Increasing the network size, the effective threshold diverges: Asymptotically, there is a giant component for any $k_a > 0$.

For $\gamma > 3$, we consider a hard cutoff $M = N^{1/(\gamma-1)}$ for the degree sequence generated in the UCM model, in order to avoid the possible appearance of outliers having a degree much larger than the average k_{max} [38]. Figures 5(c) and 5(d) show that also in this case the fraction of active nodes in the CMPGC is extensive, and its dependence on k_a is well described by Eq. (17). This result confirms the depicted scenario about the formation of an extensive CMPGC and points out that for the sizes considered only the preasymptotic scaling regime S_1 is observed, while, as expected, we do not see any trace of the asymptotic behavior (for an infinite network) $S = S_2 \approx k_a^{1-\gamma}$.
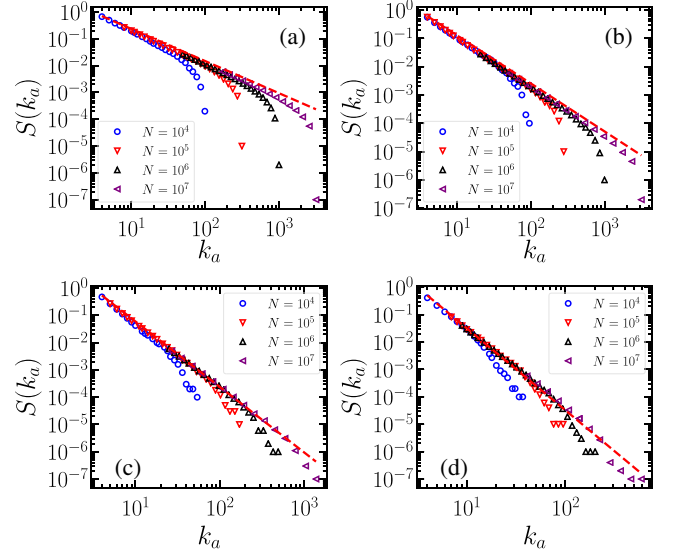

FIG. 5. Fraction S of nodes in the largest CMP component as a function of k_a for various γ values: $\gamma = 2.2$ (a), $\gamma = 2.7$ (b), $\gamma = 3.2$ (c), and $\gamma = 3.5$ (d). In all cases, $k_{min} = 3$. Symbols represent numerical results for various network sizes. Dashed lines are theoretical predictions from Eqs. (8) [(a),(b)] and (17) [(c),(d)].

Concerning finite-size effects, for $\gamma > 3$ as only the first scaling regime is observed, the condition setting the effective threshold is Eq. (24). A direct numerical verification of it for CMP is very hard, as practically all nonisolated nodes are part of the CMPGC, and finite clusters (upon which methods to determine the position of the threshold are based) are extremely rare. An indirect numerical verification is provided below in the application to the SIS model. The observation of the effective threshold associated to the second mechanism [Eq. (23)] is impossible in practice, as it requires huge networks of a size larger than N_2^*.

The conclusion of our analysis is that, in different manners depending on whether $\gamma < 3$ or $\gamma > 3$, a CMP giant component is present in infinite networks for any value of k_a. The threshold for this class of CMP processes is infinite for any value of γ.

VI. APPLICATION TO SIS EPIDEMIC SPREADING

The theoretical picture presented in the previous sections can be applied to the CMP process associated to SIS dynamics, which is an instance of this class with $k_a = a/\lambda^2 \ln(1/\lambda)$, initial mass equal to the degree, and $r(m) = m/k_a$; see Appendix A. This application has mainly the goal of investigating the properties of the SIS epidemic transition for $\gamma > 3$. We notice that the relation between CMP and SIS depends on the parameter a relating k_a with λ, whose value, either $a = 1$ or $a = 4$, is not theoretically determined. In our application of CMP to SIS, we choose to compare with both values.

A. Scaling of the CMP giant component

The scaling of S with λ for $\gamma < 3$ is obtained by inserting the expression for k_a as a function of λ into Eq. (8), obtaining

$$S = \frac{N_a}{N} \sim k_a^{1-\gamma} \sim \lambda^{2(\gamma-1)} \ln^{1-\gamma}\left(\frac{1}{\lambda}\right). \quad (25)$$

Thus, the approach predicts the existence of a CMPGC for any value of $\lambda > 0$. Notice, however, that, while it is possible to define a CMP process associated to SIS dynamics for any value of γ, the SIS epidemic transition for $\gamma < 5/2$ is due to a mechanism different from the mutual reinfection of distant hubs [21]: Hence, SIS critical properties have nothing to do with those of CMP in this case. Moreover, the connection between the scaling of S and the scaling of the SIS prevalence is not trivial in this case; hence, we cannot derive from CMP any prediction on the latter even for $5/2 < \gamma < 3$.

For $\gamma > 3$, the fraction of active nodes in the CMPGC is extensive, and its preasymptotic dependence on λ is obtained by plugging the expression for k_a into Eq. (17):

$$S_1 = \frac{N_{\mathrm{NI}}}{N} \sim k_a^{2(2-\gamma)} \sim \lambda^{4(\gamma-2)} \ln^{2(2-\gamma)}\left(\frac{1}{\lambda}\right). \quad (26)$$

We can also calculate the asymptotic scaling of the CMPGC, by plugging the expression for k_a into the expression of the scaling of the CMP giant component in the second regime, Eq. (22), obtaining

$$S_2 \sim \ln^{1-\gamma}(k_a) k_a^{1-\gamma} \sim \lambda^{2(\gamma-1)} \ln^{2(1-\gamma)}\left(\frac{1}{\lambda}\right). \quad (27)$$

We recall, however, that this scaling occurs only for exceedingly large values of k_a (i.e., values of λ exceedingly small), so that it cannot be observed in present simulations.

B. Finite-size epidemic threshold

For $\gamma > 3$, as only the first scaling regime is observed, the condition setting the effective threshold is $k_{\max}(N) = k_c(\lambda)$, i.e.,

$$\frac{a}{\lambda_c^2} \ln\left(\frac{1}{\lambda_c^2}\right) = k_{\min} k_{\max}^{1/(\gamma-2)}. \quad (28)$$

This result translates (apart from logarithmic corrections) into

$$\lambda_c(N) = (a/k_{\min})^{1/2} k_{\max}^{-1/[2(\gamma-2)]}. \quad (29)$$

For $k_{\max}^{-1/2} < \lambda < \lambda_c(N)$, there are active hubs in the system, but they do not give rise to a CMPGC. Hence, $\lambda_c(N)$ can be identified with the effective size-dependent epidemic
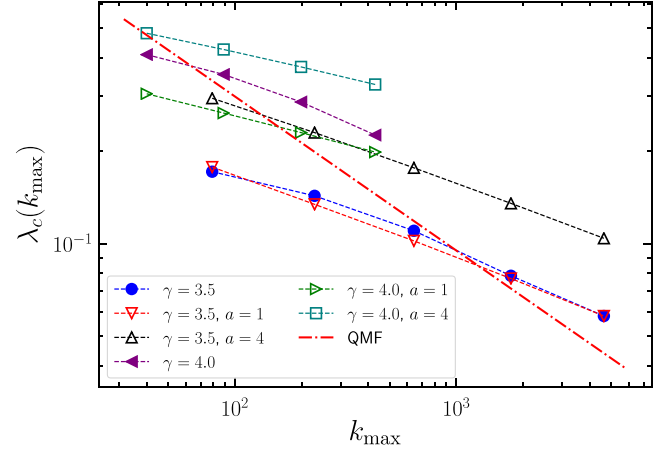

FIG. 6. Comparison between the theoretical finite-size threshold (for two different values of a) (hollow symbols) and direct numerical simulations (full symbols) of the SIS process in UCM networks with degree exponent $\gamma = 3.5$ ($k_{\min} = 3$) and $\gamma = 4$ ($k_{\min} = 2$) [24]. The epidemic threshold is determined by means of the lifespan method [24]. The dashed red line is proportional to the prediction $1/\sqrt{k_{\max}}$ of the QMF theory. The values of $k_{\max}$ correspond to sizes ranging from $N = 10^4$ to $N = 10^8$ for $\gamma = 3.5$ and from $N = 10^4$ to $N = 10^7$ for $\gamma = 4$.

threshold. Equation (29) is very interesting, as it shows that the effective threshold does not vanish as $k_{\max}^{-1/2}$, as predicted by the QMF theory, but *more slowly*, with an exponent that is reduced as γ is increased. The prediction of Eq. (28) is compared in Fig. 6 with SIS numerical results of Ref. [24], displaying a good agreement and thus clarifying a long-standing open issue. For reference, we also report the scaling predicted by the QMF theory, which patently disagrees with numerical results.

Notice, however, that this result is not the final asymptotic behavior of $\lambda_c(N)$. For much larger networks, it could be possible (at least in principle) to reach values of k_a larger than the crossover value k_2^*. In such a case, the decay of the effective threshold would be given by the condition $k_x = k_{\max}(N)$, that, from Eq. (20), leads to

$$\lambda_c(N) = \omega^{1/2} k_{\max}^{-1/2} \sim \ln(k_{\max}) k_{\max}^{-1/2}. \quad (30)$$

In this way, we recover the asymptotic scaling of the effective threshold recently derived by Huang and Durrett [27].

In Appendix D, we show that the CMP approach provides the correct effective finite-size threshold also in the case of stretched exponential degree distributions.

C. SIS prevalence as a function of λ

Above the size-dependent effective threshold, there is a backbone of active nodes which sustain an endemic state by reinfecting each other. In an infinite network, when the CMP giant component is formed by distant, mutually

interacting hubs (second regime), we can estimate the value of the prevalence (average density of infected nodes) for small λ using the following argument. All actives nodes with a degree larger than $k_x = \omega k_a$ participate in the CMPGC. Each one of these active nodes of degree k infects a number of other nodes of order λk. Since hubs are distant, these clusters of infected nodes do not overlap; hence, the total prevalence in the system is expected to be [23]

$$\rho \sim \int_{k_x}^{\infty} dk \lambda k P(k) \sim \lambda (\omega k_a)^{2-\gamma}. \qquad (31)$$

Substituting the values of ω and k_a into Eq. (31) leads to

$$\rho(\lambda) \sim \lambda^{2\gamma-3}[\ln(1/\lambda)]^{2(2-\gamma)}, \qquad (32)$$

in agreement with the exact mathematical results of Mountford, Valesin, and Yao [28]. As discussed above, this prediction is, however, impossible to verify numerically, because the onset of the asymptotic regime could be seen only for exceedingly large networks, which explains the mismatch between the theory of Mountford, Valesin, and Yao and numerical results. In doable simulations of the SIS model, the small λ regime that can be observed is the preasymptotic regime S_1 for the corresponding CMP process. In such a regime, since hubs are not well separated, it is not possible to assume that each of them independently infects a number of neighbors of the order of λk. The derivation of the exponent characterizing the SIS prevalence singularity in this preasymptotic (but long) regime remains an interesting open question for future research.

VII. DISCUSSION

In this paper, we consider a long-range percolation process, the cumulative merging percolation, exhibiting a rich phenomenology that we have uncovered developing an appropriate scaling theory. While we mainly focus on particular forms of the model inspired by the analysis of SIS process [29], more complex scenarios can be obtained by changing the functional form of the interaction range $r(m)$ or the activation probability p_i and by considering a more complicated mass merging function $m_{\alpha+\beta} = g(m_\alpha, m_\beta)$. In this sense, we expect other types of percolation transitions to arise as these features are changed. For example, if $r(m)$ saturates to a finite value when m diverges, the arguments presented above imply the presence of a finite threshold for $\gamma > 3$ as for the DOP process. The investigation of the general phenomenology of the CMP process and of its connections with other models is a promising avenue for future research.

Concerning the application of CMP to SIS dynamics, our results clarify how the mutual reinfection mechanism among distant hubs, underlying the epidemic transition for $\gamma > 3$ [24], takes place. This clarification closes the last gap in our understanding of the SIS dynamics and leads to a

HMF	ρ decays exponentially		Finite ρ
QMF	ρ decays exponentially	Finite ρ	
QMF*	ρ decays exponentially	ρ decays slowly	Finite ρ
CMP	ρ decays exponentially	ρ decays slowly	Finite ρ
	λ_c^{QMF}	λ_c	λ_c^{HMF}

FIG. 7. Behavior of SIS prevalence ρ according to the different approaches. QMF* stands for the QMF theory as reinterpreted in Refs. [22,23].

complete and consistent physical picture that we sketch here.

The original heterogeneous-mean-field theory (HMF) [15,16], based on an annealed network approximation [2], predicts an epidemic threshold given $\lambda_c^{\mathrm{HMF}} = \langle k \rangle / \langle k^2 \rangle$ and thus finite in the limit of infinite-size networks for $\gamma > 3$ (see Fig. 7). Below λ_c^{HMF}, this theory predicts a density of infected individuals $\rho(t)$ decaying exponentially to zero. For $\lambda > \lambda_c^{\mathrm{HMF}}$, the HMF predicts a finite ρ in the steady state.

The quenched-mean-field theory (QMF or NIMFA) [17–19] predicts the same scenario, a transition separating an active steady state with finite prevalence from an absorbing phase where the prevalence decays exponentially to zero. The difference with respect to the HMF is in the value of the threshold, $\lambda_c^{\mathrm{QMF}} = 1/\Lambda_M$, where Λ_M is the largest eigenvalue of the adjacency matrix, which vanishes as N diverges, for any γ. The value of the QMF threshold is the minimum value of λ such that the star graph composed by the largest hub and its direct neighbors is able to independently sustain long-lasting activity [21]. Correspondingly, the principal eigenvector for $\gamma > 5/2$ is localized around the largest hub [22]. Also, the other leading eigenvalues of the adjacency matrix are associated to eigenvectors localized on each of the hubs of the network. In the thermodynamic limit, for any given value of λ, each large hub with degree $k > 1/\lambda^2$ sustains long-lasting activity together with its direct neighbors, yielding an overall finite density of infected individuals that can be estimated [23] as $\rho \sim \lambda^{2\gamma-3}$.

However, as pointed out in Refs. [22,23], this scenario cannot really hold for SIS dynamics. In the QMF theory, there are no stochastic fluctuations, and activity in a star graph composed by $k+1$ nodes persists forever if $\lambda > 1/\sqrt{k}$. In SIS dynamics, instead, activity survives, in a star graph made of $k+1$ nodes, only for a time of the order of $\exp(\lambda^2 k/a)$. Hence, if star graphs are independent, the overall activity does not survive for a time scaling exponentially with the system size N. In other words, there is some activity surviving for some time but not a truly steady active state. In the interval $\lambda_c^{\mathrm{QMF}} < \lambda < \lambda_c^{\mathrm{HMF}}$, one

should expect a Griffiths-like phase, with a slow decay of $\rho(t)$ [23], due to the convolution of contributions from a decreasing number of still active individual hubs. Only above a *finite* threshold, approximately equal to λ_c^{HMF}, corresponding to the inverse of the first eigenvalue associated to a delocalized eigenvector [22], is a truly active steady state expected (see Fig. 7). This scenario is consistent, but at odds with numerical simulations [24] and exact mathematical results [39,40], which find an active steady state for any value of $\lambda > 0$ in the thermodynamic limit.

One possible way to reconcile these findings was explored by Lee, Shim, and Noh [23]. If large hubs are in mutual direct contact (i.e., they form an extensive connected cluster), should activity spontaneously disappear in one of them, neighboring hubs would be able to reinfect it; these mutual reinfections would lead to a survival time exponential in N, i.e., a truly active steady state. Unfortunately, the study of degree-ordered-percolation reveals that such an extensive cluster of large hubs always exists only for $\gamma < 3$. For $\gamma > 3$, the DOP threshold is finite: The largest hubs are separated and do not form an extensive cluster [23].

Our results go beyond those of Lee, Shim, and Noh and clarify what is missing in previous approaches: The activity triggered by hubs extends beyond nearest neighbors, up to a scale that grows with λ, so that hubs can interact even if they are not in direct contact. This long-range interaction gives rise, above a critical value $\lambda_c(N)$, to an extensive CMP percolating cluster of active nodes able to reinfect each other at distance and, thus, giving a veritable steady state with finite prevalence ρ. The threshold $\lambda_c(N)$ is intermediate between λ_c^{QMF} and λ_c^{HMF} and vanishes as a function of N (at odds with λ_c^{HMF}) but more slowly than λ_c^{QMF}. Considering finite networks, while for $\lambda_c(N) < \lambda < \lambda_c^{\mathrm{HMF}}$ a CMP percolating cluster exists and prevalence is finite, for $\lambda_c^{\mathrm{QMF}} < \lambda < \lambda_c$ only small nonpercolating CMP clusters are present. In this case, each of them decays independently, and, thus, a Griffiths-like phase, characterized by $\rho(t)$ slowly decaying to zero, is expected (Fig. 7). A numerical validation of this prediction, which is difficult as both interval bounds vanish with the system size, remains a challenge for future numerical studies.

The consideration of long-range effects is the crucial ingredient in our analysis that makes a qualitative difference with previous approaches. While the QMF theory neglects correlations among the dynamical state of neighbors, other theories [41–43] take some correlations into account, but, since they consider only neighbors in a short range, they cannot capture the long-range percolative nature of the SIS epidemic transition for $\gamma > 3$.

Our work puts in proper place the different theories presented in recent years to explain the behavior of the SIS model in heterogeneous networks, showing, in particular, the limit in which exact mathematical results are expected to be observed, putting thus an end to the long debate on this subject. On the other hand, it opens new perspectives, as it proposes the cumulative merging of distant clusters as a very generic phenomenon which may originate nontrivial types of percolation phenomena in networks.

APPENDIX A: CONNECTION BETWEEN CMP AND SIS

The SIS model, often called the contact process in the community of applied probabilists, is defined as follows: Individuals can be in one of two states, either susceptible or infected. Susceptible individuals become infected by contact with infected individuals, at a rate equal to the number of infected contacts times a given spreading rate β. Infected individuals, on the other hand, become spontaneously healthy again at a rate μ. The ratio $\lambda = \beta/\mu$ is the control parameter for the model, which experiences a transition between a healthy and an endemic (infected) steady state when λ crosses an epidemic threshold λ_c. In power-law distributed networks, for $\gamma > 5/2$ the epidemic transition is triggered by nodes with a large number k of neighbors (hubs). Each of these hubs together with its direct neighbors (leaves) forms a star graph, which in isolation is able to sustain the survival of the epidemic for a long time, $\tau(k) \sim \exp(\lambda^2 k/4)$ [24], provided λ is larger than $\lambda_c(k) = 1/\sqrt{k}$. During this long time interval, even if the hub recovers from the infection, it is promptly reinfected by one of its neighbors and can in its turn reinfect other leaves when they recover. After a typical time τ, a fluctuation takes the star graph formed by a hub and its nearest neighbors to the absorbing state.

Since the star graph is not isolated in the network, it can propagate activity to other nodes. It is possible to estimate [24] the average time it takes for an infected node to infect for the first time a node at distance r in the limit of small λ:

$$T(r) \sim e^{r\ln(1/\lambda)}. \qquad (A1)$$

By equating τ and $T(r)$, it is possible to estimate the "range of interaction" of a hub of degree k, i.e., the maximum distance at which a star surrounding an active hub is able to propagate the infection before spontaneously recovering:

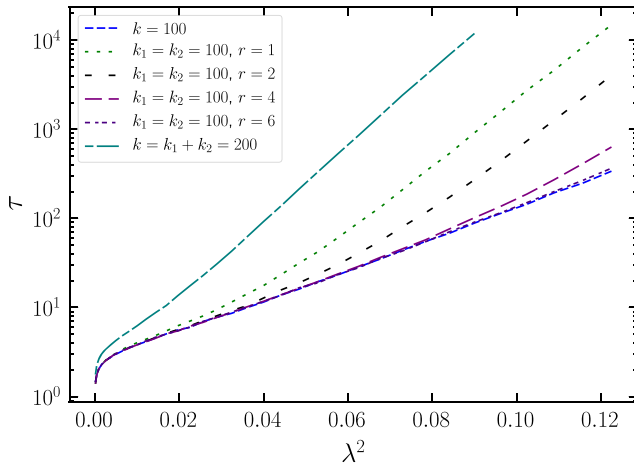$$r(k) = \frac{\lambda^2 k}{4\ln(1/\lambda)} = \frac{k}{k_a}, \qquad (A2)$$

where we define

FIG. 8. Average survival time τ of a SIS epidemic on two star graphs of size $k_1 = k_2 = k$, connected by a line of $r - 1$ intermediate nodes, starting with only the hub of one of them in the infected state. We compare with the result for single stars of size k and $2k$. As we can see, for sufficiently small values of r (distance between the hubs), the survival time of the connected stars of size k scales with λ^2 as a single star of size $2k$.

$$k_a = 4 \frac{1}{\lambda^2} \ln\left(\frac{1}{\lambda}\right). \qquad (A3)$$

Consider now another hub, of degree k' at a distance r_0 from the first. If $r_0 < r(k)$, the second hub is infected by the first and it is able to stay infected (together with its direct neighbors) for a time $\tau(k')$. During this time interval, it spreads the infection up to a distance $r(k')$. If $r_0 > r(k')$, this means that the second hub is not able to reinfect the first, should it fall into the absorbing state. Conversely, if $r_0 < r(k')$, even if the first hub recovers, it is reinfected by the second. In this way, the two distant hubs form a coupled system such that if one hub recovers, the other is able to reinfect it before recovering in its turn. For the infection to die out in the system of the two hubs, they must recover almost simultaneously [29]. These concurrent recoveries happen after a time of the order of $\tau(k)\tau(k') \sim \exp[\lambda^2(k + k')]$; see Fig. 8.

Hence, the combined set of hubs are able to infect nodes at an increased range of interaction $r(k + k')$. It is then clear that SIS dynamics can be seen as an instance of the cumulative merging process, with active nodes those with $k \geq k_a = 4/\lambda^2 \ln(1/\lambda)$, initial masses equal to node degrees $m_i^{(0)} = k_i$, and range of interaction given by $r(m) = m/k_a$. Notice that the factor of 4 in the expression for k_a is the consequence of the choice $\tau(k) \sim \exp(\lambda^2 k/4)$. Alternative treatments [27,44] give that star graphs are active for $k > k_a = 1/\lambda^2 \ln(1/\lambda)$. In the comparison of the CMP approach to SIS with numerical simulations, we consider both expressions.

APPENDIX B: AVERAGE DISTANCE BETWEEN A NODE OF DEGREE k AND THE CLOSEST NODE OF DEGREE AT LEAST k

We can estimate the average distance $d(k)$ between a node of degree k and the nearest node of degree larger than or equal to k within a treelike approximation [2]. For random uncorrelated networks, the probability that a link points to a node of degree k' is $k'P(k')/\langle k \rangle$. Arriving at this node, there are $k' - 1$ possible outgoing edges (excluding the one used to arrive to node k'). The average number of outgoing edges (the so-called branching factor) is, thus,

$$\kappa = \int_{k_{\min}}^{\infty} (k' - 1) \frac{k'P(k')}{\langle k \rangle} dk' = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1, \qquad (B1)$$

that is, a finite number for power-law networks with $\gamma > 3$. From this branching ratio, we estimate the average number of nodes at distance n as $N_n = k\kappa^{n-1}$, assuming the tree approximation.

A node of degree k has k neighbors. It is connected at distance $d = 1$ to a node of degree not less than k if at least one of these neighbors has a degree larger than or equal to k. The probability of this event is

$$P_>(k) = \int_k^{\infty} \frac{k'P(k')}{\langle k \rangle} dk' = \left(\frac{k}{k_{\min}}\right)^{2-\gamma}. \qquad (B2)$$

Therefore, the probability that the distance at the nearest neighbors with a degree larger than or equal to k is equal to $d = 1$ is

$$P(d = 1) = 1 - [1 - P_>(k)]^k = 1 - [P_<(k)]^k, \qquad (B3)$$

where $P_<(k) = 1 - P_>(k)$ is the probability that a nearest neighbor of a node has a degree smaller than k.

The nearest neighbor with a degree not less than k is at distance $d = 2$ if there are no such neighbors at distance $d = 1$, and at least one of the neighbors at distance $d = 2$, in number $k\kappa$, has a degree not less than k, which happens with probability

$$P(d = 2) = [P_<(k)]^k \{1 - [P_<(k)]^{k\kappa}\}. \qquad (B4)$$

By induction, we can see that the nearest neighbor of a degree not less that k is at distance $d = n$ corresponds to not observing one at any distance smaller than n and having at least one at a distance equal to n, which happens with probability

$$P(d = n) = [P_<(k)]^k [P_<(k)]^{k\kappa} [P_<(k)]^{k\kappa^2} \dots [P_<(k)]^{k\kappa^{n-2}} \{1 - [P_<(k)]^{k\kappa^{n-1}}\}$$

$$= [P_<(k)]^{\sum_{r=0}^{n-2} k\kappa^r} - [P_<(k)]^{\sum_{r=0}^{n-1} k\kappa^r}$$

$$= [P_<(k)]^{k(\kappa^{n-1}-1)/(\kappa-1)} - [P_<(k)]^{k(\kappa^n-1)/(\kappa-1)}$$

$$= \frac{[P_<(k)]^{k\kappa^{n-1}/(\kappa-1)} - [P_<(k)]^{k\kappa^n/(\kappa-1)}}{[P_<(k)]^{k/(\kappa-1)}} \equiv \frac{C^{\kappa^{n-1}} - C^{\kappa^n}}{C}, \tag{B5}$$

where for simplicity we set $C = [P_<(k)]^{k/(\kappa-1)}$.

The average distance $d(k)$ can be evaluated as

$$d(k) = \sum_{n=1}^{\infty} nP(d = n) = \sum_{n=1}^{\infty} n\frac{C^{\kappa^{n-1}} - C^{\kappa^n}}{C}$$

$$= \sum_{n=0}^{\infty} \frac{C^{\kappa^n}}{C}. \tag{B6}$$

The summation in Eq. (B6) cannot be performed directly. We can approximate its behavior for large k by transforming it into an integral:

$$d(k) \simeq \frac{1}{C} \int_0^{\infty} C^{\kappa^x} dx = \frac{1}{C \ln(\kappa)} \int_1^{\infty} \frac{C^z}{z} dz$$

$$= \frac{\Gamma[0, -\ln(C)]}{C \ln(\kappa)}, \tag{B7}$$

where $\Gamma(a, z)$ is the incomplete Gamma function [36] and we apply the change of variables $\kappa^x = z$. For large k, $C =$
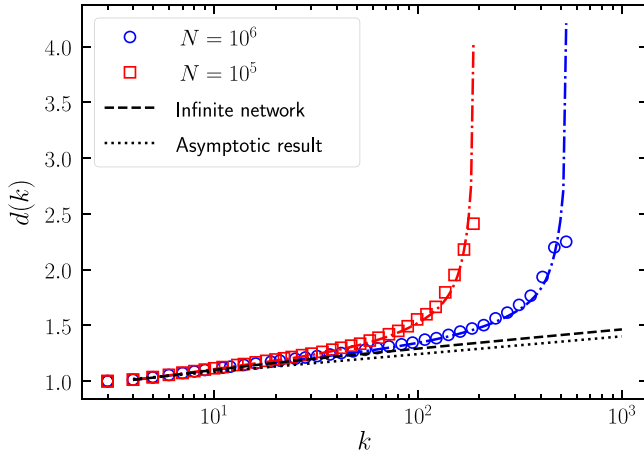

FIG. 9. Average distance between a node of degree k and the closest node of a degree at least k in power-law networks with degree exponent $\gamma = 3.2$. Symbols represent numerical simulations over networks of different sizes, averaged over 10^4 independent network samples. The dashed line represents the numerical evaluation of the summation Eq. (B6) in the infinite network limit. Dot-dashed lines represent results of the summation for finite networks of the corresponding size N. The dotted line corresponds to the asymptotic expression Eq. (B9).

$[1 - (k/k_{\min})^{2-\gamma}]^{k/(\kappa-1)}$ tends to 1, so we can expand the incomplete Gamma function in Eq. (B8) for small arguments, $\Gamma(0, z) \sim -\ln(z)$ [36], to obtain the asymptotic behavior

$$d(k) \sim \frac{-\ln[-\ln(C)]}{C \ln(\kappa)} \sim \frac{\ln\left[(\frac{k}{k_{\min}})^{\gamma-3} \frac{\kappa-1}{k_{\min}}\right]}{\ln(\kappa)}, \tag{B8}$$

where we expand C for large k. We therefore observe the asymptotic behavior for large k in infinite networks as

$$d(k) \sim 1 + \frac{\gamma - 3}{\ln(\kappa)} \ln\left(\frac{k}{k_{\min}}\right), \tag{B9}$$

where the term 1 accounts for the minimum possible distance between nodes.

This calculation, performed in the tree approximation, captures nevertheless the behavior in real uncorrelated power-law networks generated with the UCM [35]. In Fig. 9, we present the result of numerical simulations, together with the numerical evaluation of the summation in Eq. (B6), performed using a discrete power-law degree distribution $P(k) = k^{-\gamma}/[\zeta(\gamma, k_{\min}) - \zeta(\gamma, k_{\max})]$, where $\zeta(s, a)$ is the Hurwitz zeta function [36]. The dashed line represents the result for an infinite network ($k_{\max} = \infty$), while the dot-dashed lines mark the value for networks with maximum degree $k_{\max} = N^{1/(\gamma-1)}$ [37]. The dotted line shows the asymptotic behavior obtained in Eq. (B9).

APPENDIX C: CUMULATIVE MERGING PERCOLATION ON STRETCHED EXPONENTIAL NETWORKS

Let us consider the example of a network with cumulative degree distribution [27]

$$P_c(k) = e^{-k^\beta + k_{\min}^\beta}, \tag{C1}$$

corresponding to a stretched exponential degree distribution

$$P(k) = -\frac{dP_k(k)}{dk} = \beta k^{\beta-1} e^{k_{\min}^\beta - k^\beta}. \tag{C2}$$

Applying the extreme value theory, for a finite network of size N we have

$$k_{\max} \sim [\ln(N)]^{1/\beta}. \tag{C3}$$

The other relevant quantities for CMP are

$$\frac{N_a}{N} = \int_{k_a}^{\infty} dk P(k) = e^{k_{\min}^{\beta} - k_a^{\beta}} \tag{C4}$$

and

$$P_a = \frac{1}{k_c} = \int_{k_a}^{\infty} dk \frac{kP(k)}{\langle k \rangle} = \frac{\Gamma(1 + \frac{1}{\beta}, k_a^{\beta})}{\Gamma(1 + \frac{1}{\beta}, k_{\min}^{\beta})} \tag{C5}$$

$$\simeq \frac{e^{-k_a^{\beta}} k_a}{\Gamma(1 + \frac{1}{\beta}, k_{\min}^{\beta})}, \tag{C6}$$

where we develop the numerator in the limit of large k_a. The average number of active neighbors for each active node is

$$\frac{N}{N_a} \int_{k_a}^{\infty} dk k P(k) P_a = \frac{e^{k_a^{\beta}} \Gamma(1 + \frac{1}{\beta}, k_a^{\beta})^2}{\Gamma(1 + \frac{1}{\beta}, k_{\min}^{\beta})}$$

$$\sim \frac{e^{-k_a^{\beta}} k_a^2}{\Gamma(1 + \frac{1}{\beta}, k_{\min}^{\beta})}, \tag{C7}$$

where we expand the last expression in the limit of large k_a. Therefore, the average number of active neighbors of an active node vanishes exponentially. This result implies that the size of the DOPGC decays exponentially fast and the extended DOP mechanism is not at work: Small clusters are at a distance much larger than 2 from the DOPGC.

The only mechanism leading to the formation of the CMPGC is the second one, based on the interaction at a distance among isolated nodes. This interaction involves a scale $k_x = \omega k_a$, such that $r(k_x) = d(k_x)$, to ensure that all nodes with $k > k_x$ see each other and can merge in the same cluster. To compute $d(k)$, from Appendix B we must evaluate, in the limit of large k,

$$d(k) \sim \frac{-\ln[-\ln(C)]}{C \ln(\kappa)}, \tag{C8}$$

with $C = [P_<(k)]^{k/(\kappa-1)}$ and κ the branching factor. In this case,

$$P_<(k) = 1 - P_c(k) = 1 - e^{k_{\min}^{\beta} - k^{\beta}}. \tag{C9}$$

For large k,

$$-\ln(C) \simeq \frac{k}{\kappa - 1} e^{k_{\min}^{\beta} - k^{\beta}}, \tag{C10}$$

and

$$-\ln[-\ln(C)] \simeq k^{\beta} - k_{\min}^{\beta} - \ln\left(\frac{k}{\kappa - 1}\right). \tag{C11}$$

Therefore, for large k,

$$d(k) \simeq \frac{k^{\beta}}{\ln(\kappa)}, \tag{C12}$$

where we disregard constant and logarithmic terms. For $r(k) = k/k_a$, from $d(k_x) = r(k_x)$, we obtain

$$\omega = \frac{\omega^{\beta} k_a^{\beta}}{\ln(\kappa)}, \tag{C13}$$

leading to

$$\omega = \left(\frac{k_a^{\beta}}{\ln(\kappa)}\right)^{1/(1-\beta)}. \tag{C14}$$

So, we have

$$k_x = k_a \omega = \left(\frac{k_a}{\ln(\kappa)}\right)^{1/(1-\beta)}. \tag{C15}$$

As a consequence,

$$S_2 = \int_{k_x}^{\infty} dk P(k) = e^{k_{\min}^{\beta} - k_x^{\beta}} \tag{C16}$$

$$\approx e^{k_{\min}^{\beta} - [k_a/\ln(\kappa)]^{\beta/(1-\beta)}}. \tag{C17}$$

APPENDIX D: APPLICATION TO SIS ON STRETCHED EXPONENTIAL NETWORKS

Since the asymptotic behavior of the order parameter for the CMP transition is given by S_2, the effective finite-size threshold is given by the condition $k_x \simeq k_{\max}$, that is,

$$k_a \simeq \ln(\kappa) k_{\max}^{1-\beta}. \tag{D1}$$

Using $k_a = a(1/\lambda)^2 \ln(1/\lambda)$, this equation implies

$$a(1/\lambda_c)^2 \ln(1/\lambda_c) \simeq \ln(\kappa) k_{\max}^{1-\beta}. \tag{D2}$$

Disregarding logarithmic factors, this expression can be inverted, leading, in the limit of large $k_{\max}$, to

$$\lambda_c \simeq \sqrt{\frac{a(1-\beta)}{2\ln(\kappa)}} [k_{\max}^{\beta-1} \ln(k_{\max})]^{1/2}. \tag{D3}$$

For a stretched exponential degree distribution, $k_{\max} \simeq [\ln(N)]^{1/\beta}$, so we finally have

$$\lambda_c \simeq \sqrt{\frac{a(1-\beta)}{2\beta\ln(\kappa)}}[\ln(N)]^{(\beta-1)/(2\beta)}\{\ln[\ln(N)]\}^{1/2}. \quad (D4)$$

In this way, we recover the exact logarithmic dependence of the effective threshold on N for the stretched exponential case, recently found in Ref. [27].

In the limit of a pure exponential distribution, $\beta = 1$, the previous arguments cannot be applied. However, recent results in Ref. [45] show that the threshold in this case is finite.

---

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

[2] S. N. Dorogovtsev, A. V. Goltsev, and J F. F. Mendes, *Critical Phenomena in Complex Networks*, Rev. Mod. Phys. **80,** 1275 (2008).

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic Processes in Complex Networks*, Rev. Mod. Phys. **87,** 925 (2015).

[4] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Vol. 4 (Springer, New York, 2016).

[5] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, Interdisciplinary Applied Mathematics Vol. 46 (Springer, New York, 2017).

[6] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley, New York, 2000).

[7] P. Grassberger, *On the Critical Behavior of the General Epidemic Process and Dynamical Percolation*, Math. Biosci. **63,** 157 (1983).

[8] M. Newman, *Spread of Epidemic Disease on Networks*, Phys. Rev. E **66,** 016128 (2002).

[9] Bond percolation is exactly mapped to the final state of the independent cascade model, a SIR-like process having a fixed recovery time. For SIR, the mapping is on a semidirected epidemic percolation network, as explained in Ref. [10].

[10] E. Kenah and J. Robins, *Second Look at the Spread of Epidemics on Networks*, Phys. Rev. E **76,** 036113 (2007).

[11] M. Á. Serrano and M. Boguñá, *Percolation and Epidemic Thresholds in Clustered Networks*, Phys. Rev. Lett. **97,** 088701 (2006).

[12] B. Karrer and M. E. J. Newman, *Message Passing Approach for General Epidemic Models*, Phys. Rev. E **82,** 016101 (2010).

[13] J. P. Gleeson and M. A. Porter, *Message-Passing Methods for Complex Contagions*, in *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks*, edited by S. Lehmann and Y.-Y. Ahn (Springer, New York, 2018), pp. 81–95.

[14] A.-L. Barabási and R. Albert, *Emergence of Scaling in Random Networks*, Science **286,** 509 (1999).

[15] R. Pastor-Satorras and A. Vespignani, *Epidemic Spreading in Scale-Free Networks*, Phys. Rev. Lett. **86,** 3200 (2001).

[16] R. Pastor-Satorras and A. Vespignani, *Epidemic Dynamics and Endemic States in Complex Networks*, Phys. Rev. E **63,** 066117 (2001).

[17] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS'03), Los Alamitos, CA* (IEEE, New York, 2003), pp. 25–34.

[18] P. Van Mieghem, J. Omic, and R. E. Kooij, *Virus Spread in Networks*, IEEE/ACM Trans. Netw. **17,** 1 (2009).

[19] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Discrete-Time Markov Chain Approach to Contact-Based Disease Spreading in Complex Networks*, Europhys. Lett. **89,** 38009 (2010).

[20] C. Castellano and R. Pastor-Satorras, *Thresholds for Epidemic Spreading in Networks*, Phys. Rev. Lett. **105,** 218701 (2010).

[21] C. Castellano and R. Pastor-Satorras, *Competing Activation Mechanisms in Epidemics on Networks*, Sci. Rep. **2,** 371 (2012).

[22] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, *Localization and Spreading of Diseases in Complex Networks*, Phys. Rev. Lett. **109,** 128702 (2012).

[23] H. K. Lee, P.-S. Shim, and J. D. Noh, *Epidemic Threshold of the Susceptible-Infected-Susceptible Model on Complex Networks*, Phys. Rev. E **87,** 062812 (2013).

[24] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Nature of the Epidemic Threshold for the Susceptible-Infected-Susceptible Dynamics in Networks*, Phys. Rev. Lett. **111,** 068701 (2013).

[25] A. S. Mata, M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Lifespan Method as a Tool to Study Criticality in Absorbing-State Phase Transitions*, Phys. Rev. E **91,** 052117 (2015).

[26] G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, *Fundamentals of Spreading Processes in Single and Multilayer Complex Networks*, Phys. Rep. **756,** 1 (2018).

[27] X. Huang and R. Durrett, *The Contact Process on Random Graphs and Galton Watson Trees*, ALEA, Lat. Am. J. Probab. Math. Stat. **17,** 159 (2020).

[28] T. Mountford, D. Valesin, and Q. Yao, *Metastable Densities for the Contact Process on Power Law Random Graphs*, Electron. J. Pro **18,** 36 (2013).

[29] L. Ménard and A. Singh, *Percolation by Cumulative Merging and Phase Transition for the Contact Process on Random Graphs*, Ann. Sci. École Norm. Sup. **49,** 1189 (2016).

[30] G. Grimmett, *Percolation*, 2nd ed. (Springer-Verlag, Berlin, 1999).

[31] In Ref. [29], it is demonstrated that the CMP threshold is a lower bound for the epidemic threshold. Based on the physical picture, we expect the two quantities to coincide.

[32] Although this result is mathematically proved only in a specific case in Ref. [29], we find numerically the same independence in the cases considered below.

[33] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, *Stability and Topology of Scale-Free Networks under Attack and Defense Strategies*, Phys. Rev. Lett. **94,** 188701 (2005).

[34] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Dynamical and Correlation Properties of the Internet*, Phys. Rev. Lett. **87**, 258701 (2001).

[35] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, *Generation of Uncorrelated Random Scale-Free Networks*, Phys. Rev. E **71**, 027103 (2005).

[36] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[37] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *Cut-offs and Finite Size Effects in Scale-Free Networks*, Eur. Phys. J. B **38**, 205 (2004).

[38] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Langevin Approach for the Dynamics of the Contact Process on Annealed Scale-Free Networks*, Phys. Rev. E **79**, 036110 (2009).

[39] S. Chatterjee and R. Durrett, *Contact Process on Random Graphs with Degree Power Law Distribution Have Critical Value Zero*, Ann. Probab. **37**, 2332 (2009).

[40] T. Mountford, J.-C. Mourrat, D. Valesin, and Q. Yao, *Exponential Extinction Time of the Contact Process on Finite Graphs*, Stoch. Proc. Appl. **126**, 1974 (2016).

[41] A. S. Mata and S. C. Ferreira, *Pair Quenched Mean-Field Theory for the Susceptible-Infected-Susceptible Model on Complex Networks*, Europhys. Lett. **103**, 48003 (2013).

[42] J. P. Gleeson, *High-Accuracy Approximation of Binary-State Dynamics on Networks*, Phys. Rev. Lett. **107**, 068701 (2011).

[43] G. St-Onge, J.-G. Young, E. Laurence, C. Murphy, and L. J. Dubé, *Phase Transition of the Susceptible-Infected-Susceptible Dynamics on Time-Varying Configuration Model Networks*, Phys. Rev. E **97**, 022305 (2018).

[44] E. Cator and P. Van Mieghem, *Susceptible-Infected-Susceptible Epidemics on the Complete Graph and the Star Graph: Exact Analysis*, Phys. Rev. E **87**, 012811 (2013).

[45] S. Bhamidi, D. Nam, O. Nguyen, and A. Sly, *Survival and Extinction of Epidemics on Random Graphs with General Degrees*, arXiv:1902.03263.