Influential spreaders for recurrent epidemics on networks

Gaël Poux-Médard^{1,2,3}, Romualdo Pastor-Satorras,⁴ and Claudio Castellano²

¹*Physics Department, École Normale Supérieure de Lyon, 15 parvis René Descartes, 69342 Lyon, France*

²*Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, 00185 Roma, Italy*

³*Laboratoire ERIC, Université Lumière Lyon 2, 5 avenue Pierre Mendès France, 69676 Bron, France*

⁴*Departament de Física, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*

(Received 13 December 2019; accepted 29 May 2020; published 12 June 2020)

The identification of which nodes are optimal seeds for spreading processes on a network is a nontrivial problem that has attracted much interest recently. While activity has mostly focused on the nonrecurrent type of dynamics, here we consider the problem for the susceptible-infected-susceptible (SIS) spreading model, where an outbreak seeded in one node can originate an infinite activity avalanche. We apply the theoretical framework for avalanches on networks proposed by D. B. Larremore *et al.* [*Phys. Rev. E* **85**, 066131 (2012)] to obtain detailed quantitative predictions for the spreading influence of individual nodes (in terms of avalanche duration and avalanche size) both above and below the epidemic threshold. When the approach is complemented with an annealed network approximation, we obtain fully analytical expressions for the observables of interest close to the transition, highlighting the role of degree centrality. A comparison of these results with numerical simulations performed on synthetic networks with power-law degree distribution reveals, in general, good agreement in the subcritical regime, leaving thus some questions open for further investigation relative to the supercritical regime.

DOI: [10.1103/PhysRevResearch.2.023332](https://doi.org/10.1103/PhysRevResearch.2.023332)

I. INTRODUCTION

Some epidemic outbreaks die out very rapidly, affecting just a few individuals, while others, caused by the same pathogen, survive for much longer times, infecting considerably larger fractions of the overall population [1]. A meme posted by a celebrity on a social network is rapidly spread to millions of users, while a similar meme posted by a less connected individual rarely becomes viral. What decides the fate of a single spreading process? One of the most important factors is where, in the connectivity pattern mediating the spreading process, the initial seed is located. Intuitively, an infection seeded in a highly connected node has a higher chance of reaching a large number of individuals, but other, less local network features may, in principle, play a role: For example, a node sitting at the boundary between two communities can be highly effective in giving rise to large outbreaks even if having only a few direct connections. The nontrivial question of how to identify “influential spreaders” in a network has been the focus of an impressive amount of work [2,3] in the last decade. In a nutshell, the problem is the prediction of how large, on average, a spreading event started by a single node in the network is.¹ A slightly more limited goal is to rank network nodes depending on their

spreading capability and to correlate such a ranking with purely topological node features, such as degree or other types of centrality.

The seminal paper by Kitsak *et al.* [4] posed in a clear manner this question, highlighting its nontriviality and reporting a strong correlation between the average size of an outbreak seeded in node n and a particular type of centrality measure, the K -core index [5]. A huge number of other centralities have been proposed and tested as ways to identify influential spreaders [3,6,7], with different levels of success, depending, in many cases, on the type of network substrate considered for the empirical validation. Among the most firmly grounded results, Ref. [8] exploited a mapping to bond percolation to relate spreading influence at the epidemic transition to non-backtracking centrality [9,10], while Ref. [11] extended the approach presenting a message-passing framework, allowing us to compute the average outbreak size for any value of the control parameter.

All these results were found for the susceptible-infected-removed (SIR) dynamics [12], the simplest model for epidemics allowing permanent immunity or simple variants of it. For this type of model the definition of spreading influence is straightforward: Throughout the whole phase space, outbreaks have a finite duration, and the average size of an outbreak seeded in a node is the measure of how influential the node is. The other large class of epidemic models, whose simplest example is the susceptible-infected-susceptible (SIS) model, describes recurrent dynamics where each node can be reinfected over and over again [12]. While some results are available [13,14], this class has received, nevertheless, much less attention. Our purpose in this paper is to fill this gap, performing a systematic study of the problem of identifying influential spreaders for a parallel, discrete-time

¹In the epidemic framework a spreading event started in a single node is called an outbreak. We will often use in the following the more general term “avalanche.”

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

version of SIS dynamics. As for nonrecurrent (SIR-like) epidemic models, the phase diagram for SIS dynamics consists of a healthy phase for small values of the control parameter (see below) separated from an epidemic phase by a critical value of the control parameter, the epidemic threshold. While the nature of the healthy phase is the same for both classes of models, the possibility of multiple reinfections changes the nature of the epidemic phase. Above the threshold there is a nonvanishing chance that a stationary “endemic” state is reached, where the epidemic survives forever, with a steady finite density of infected nodes. This possibility implies that for SIS dynamics there are different possible definitions of what an influential spreader is. Below the critical point, since all avalanches eventually end, one can take as a measure of spreading influence of node n the total duration or the total size of an avalanche seeded in n , just as in the SIR case. In the supercritical region instead, some avalanches are infinite; that is, they give rise to the never-ending (at least in the thermodynamic limit) endemic state. In this case being an influential spreader may have two distinct meanings. Good spreaders can be nodes giving rise to an infinite avalanche with large probability or nodes giving rise to large or long *finite* avalanches. These definitions of influential spreaders do not necessarily coincide. For this reason we consider three main observables of interest [15]:

(i) The first is the probability b_n that an avalanche seeded in node n is finite (i.e., it does not lead to the endemic state above the transition). Below the transition, $b_n = 1$; above it, $b_n < 1$.

(ii) The second is the average duration T_n of (finite) avalanches seeded in node n . In the following we will consider a discrete-time dynamics; hence, the duration of each avalanche will be an integer.

(iii) The third is the average size S_n of (finite) avalanches seeded in node n . This quantity is equal to the total number of activation events. Since each node can be activated more than once, S can, in principle, be larger than the total number of nodes N .

Another observable of interest is the coverage, the average number of distinct nodes reached by an avalanche [16]. We performed some comparisons between this quantity and S_n , finding minimal differences. For this reason we will not consider the coverage in the following.

II. THEORETICAL APPROACH

A. Quenched mean-field theory

The identification of influential spreaders for SIS dynamics is greatly helped by the application of the general quenched mean-field (QMF) [17–20] theoretical approach for avalanches on networks proposed by Larremore *et al.* [15]. In that paper a comprehensive theory was presented for the statistical properties of avalanches when the dynamical process is determined by a matrix A_{nm} , whose individual entry is the probability that the avalanche propagates from node n to node m . The approach predicts a transition between a subcritical regime with all avalanches finite ($b_n = 1$) and a supercritical regime where a fraction $1 - b_n$ of all avalanches lasts forever. The transition depends on the largest eigenvalue λ of the A

matrix. If $\lambda < \lambda_c^{\text{QMF}} = 1$, the system is subcritical; if $\lambda > \lambda_c^{\text{QMF}}$, the system is supercritical, and infinite avalanches start to appear.

We can apply this theory to the SIS dynamics by considering a parallel discrete-time version of it, analogous to the independent cascade model version of SIR dynamics [21]. In this dynamics, at each time step each infected node attempts to transmit the infection to all its direct neighbors, and each independent attempt succeeds with probability p . After attempting to infect all its neighbors, an infected node recovers and switches back to the susceptible state with a recovery time fixed to 1. The probability that an avalanche propagates from node n to m is therefore $A_{nm} = pa_{nm}$, where a_{nm} is the unweighted adjacency matrix of the network. Clearly, the largest eigenvalue of A is $\lambda = p\Lambda_M$, where Λ_M is the largest eigenvalue of the adjacency matrix a . As a consequence, the epidemic threshold is $p_c^{\text{QMF}} = 1/\Lambda_M$.

1. Probability of observing a finite avalanche

Within QMF theory the probability b_n of having a finite avalanche is obtained by solving iteratively the equation

$$\begin{aligned} b_n &= \prod_{m=1}^N [(1 - A_{nm}) + A_{nm}b_m] \\ &= \prod_{m=1}^N [1 - pa_{nm}(1 - b_m)]. \end{aligned} \quad (1)$$

As shown in Ref. [15], this equation admits a solution $b_n = 1$ for any λ and for $\lambda > 1$, i.e., in the supercritical regime, another stable solution $b_n < 1$.

2. Average avalanche duration

Let us consider $c_n(t)$ to be the probability that an avalanche has a duration smaller than or equal to t , with $c_n(0) = 0$, by definition. The probability $c_n(t)$ fulfills the recurrence equation [15]

$$c_n(t+1) = \prod_{m=1}^N [(1 - A_{nm}) + A_{nm}c_m(t)]. \quad (2)$$

In the limit $t \rightarrow \infty$, $c_n(t) \rightarrow b_n$, where $b_n = 1$ in the subcritical phase and $b_n < 1$ in the supercritical phase, hence Eq. (1). Consider $c_n(t) = b_n - f_n(t)$, with $f_n(0) = b_n$. For sufficiently large t , $f_n(t)$ is small, so we can linearize Eq. (2) to obtain [15]

$$f_n(t) = \sum_m D_{nm} f_m(t-1), \quad (3)$$

where for the SIS process

$$D_{nm} = \frac{A_{nm}b_n}{(1 - A_{nm}) + A_{nm}b_m} = \frac{pb_n a_{nm}}{1 - p + pb_m}, \quad (4)$$

which, in the subcritical phase, where $b_n = 1$, reduces to $D_{nm} = A_{nm} = pa_{nm}$. The avalanche duration distribution $p_n(t)$ is given by

$$p_n(t) = c_n(t) - c_n(t-1) = f_n(t-1) - f_n(t). \quad (5)$$

Therefore, we can write the average avalanche duration T_n as

$$\begin{aligned}
 T_n &= \sum_{t=1}^{\infty} t p_n(t) = \sum_{t=1}^{\infty} t f_n(t-1) - \sum_{t=1}^{\infty} t f_n(t) \\
 &= f_n(0) + \sum_{t=2}^{\infty} t f_n(t-1) - \sum_{t=1}^{\infty} t f_n(t) \\
 &= b_n + \sum_{t=1}^{\infty} (t+1) f_n(t) - \sum_{t=1}^{\infty} t f_n(t) \\
 &= b_n + \sum_{t=1}^{\infty} f_n(t) \equiv b_n + R_n.
 \end{aligned} \quad (6)$$

If $f_n(t)$ decreases sufficiently fast, we can assume that the linearization of $f_n(t)$ is valid for all times t to obtain

$$\begin{aligned}
 R_n &= \sum_{t=1}^{\infty} f_n(t) = \sum_m D_{nm} \sum_{t=1}^{\infty} f_m(t-1) \\
 &= \sum_m D_{nm} \left[f_m(0) + \sum_{t=1}^{\infty} f_m(t) \right] \\
 &= \sum_m D_{nm} [b_m + R_m].
 \end{aligned} \quad (7)$$

Combining the previous equations, we have

$$T_n = b_n + \sum_m D_{nm} T_m. \quad (8)$$

We can obtain an exact expression at the QMF level by inverting this linear relationship, obtaining

$$T = [I - D]^{-1} B, \quad (9)$$

where $T = (T_1, \dots, T_N)^T$, $B = (b_1, \dots, b_N)^T$, and I is the identity matrix.

3. Average avalanche size

From Ref. [15], the generating function of the avalanche size distribution $p_n(s)$,

$$\phi_n(z) = \sum_s e^{-sz} p_n(s), \quad (10)$$

fulfills the equation

$$\phi_n(z) = e^{-z} \prod_m [1 + (\phi_m(z) - 1) H_{nm}], \quad (11)$$

where

$$H_{nm} = \frac{b_m p a_{nm}}{1 - p + b_m p} \quad (12)$$

in the SIS process. Taking logarithms of both sides of Eq. (11), deriving with respect to z , and setting $z = 0$, one obtains

$$\frac{\phi'_n(0)}{\phi_n(0)} = -1 + \sum_m \frac{H_{nm} \phi'_m(0)}{1 + (\phi_m(0) - 1) H_{nm}}. \quad (13)$$

From the definition of the generating function we have $\phi_n(0) = 1$, while $\phi'_n(0)$ is minus the average avalanche

size,

$$\phi'_n(0) = \left. \frac{d\phi_n(z)}{dz} \right|_{z=0} = - \sum_s s p_n(s) = -S_n. \quad (14)$$

Hence, Eq. (13) reads

$$S_n = 1 + \sum_m H_{nm} S_m. \quad (15)$$

Inverting this linear relationship in matrix format, we can write the QMF solution

$$S = [I - H]^{-1} \mathbf{1}_N, \quad (16)$$

where $S = (S_1, S_2, \dots, S_N)^T$ and $\mathbf{1}_N$ is a column vector of ones, of size N . Notice that, contrary to the case of T_n , no long-time assumption has been made to derive Eq. (15).

It is interesting to observe that, due to the similarity relation $H_{nm} = \frac{b_m}{b_n} D_{nm}$ [15], the average size and duration at the QMF level are not independent but related by the identity

$$T_n = b_n S_n. \quad (17)$$

Below the threshold, when $b_n = 1$, average time and duration are equal. This fact can be interpreted in terms of avalanches that infect just a new node, on average, every time step. On the contrary, in the supercritical regime, $b_n < 1$ in general, and thus, $S_n > T_n$ on average. We remark also that, so far, we have not made any assumption about the value of p ; hence, these predictions are valid throughout the whole phase diagram.

B. Annealed network approximation

Equations (1), (8), and (15) constitute sets of N closed equations, whose straightforward solutions provide predictions about the ability of each node in a generic network to spread influence.

It is possible to obtain fully analytical results (and thus obtain more physical insight) by implementing an additional approximation. The annealed network approximation [22] assumes that the network is fully rewired at each time step while keeping individual degrees fixed. In this way dynamical correlations among different nodes are destroyed, and the state of each node can depend only on its degree k . Mathematically, the annealed network approximation consists of replacing, for uncorrelated networks, the adjacency matrix by the probabilistic form [22]

$$a_{nm} \simeq \frac{k_n k_m}{\langle k \rangle N}. \quad (18)$$

This allows us to write

$$\sum_m a_{nm} F(k_m) \simeq k_n \frac{\langle k F \rangle}{\langle k \rangle} \quad (19)$$

for any function $F(k)$ depending on the degree.

1. Probability of observing a finite avalanche

Taking logarithms on both sides of Eq. (1), we have

$$\begin{aligned}
 \ln b_n &= \sum_{m=1}^N \ln[1 - p a_{nm} (1 - b_m)] \\
 &= -p \sum_{m=1}^N a_{nm} (1 - b_m) = -p k_n + p \sum_{m=1}^N a_{nm} b_m,
 \end{aligned}$$

where we have assumed $p(1 - b_m)$ to be small, i.e., either p is small or the system is close to criticality ($b_n \simeq 1$).

Assuming that b_n is a function of the degree k_n and using Eq. (19), we can then write

$$\ln b_n = -pk_n \left(1 - \frac{\langle bk \rangle}{\langle k \rangle} \right), \quad (20)$$

leading to

$$b_n = e^{-k_n p \theta}, \quad (21)$$

where we have defined the quantity

$$\theta = 1 - \frac{\langle bk \rangle}{\langle k \rangle}, \quad (22)$$

which depends implicitly on p , through the average $\langle bk \rangle$. We can make this dependence explicit by considering the self-consistent equation

$$\theta = 1 - \sum_k \frac{kP(k)e^{-kp\theta}}{\langle k \rangle} \equiv \Psi(\theta). \quad (23)$$

The value $\theta = 0$, corresponding to $b_n = 1$, is always a solution of Eq. (23). The condition for the existence of an additional nonzero solution is that the function $y = \Psi(\theta)$ crosses the line $y = \theta$ at a point $\theta > 0$, something that happens when $\frac{d\Psi(\theta)}{d\theta}|_{\theta=0} > 1$. This condition translates into $p > p_c^{\text{ann}}$, where the threshold

$$p_c^{\text{ann}} = \frac{\langle k \rangle}{\langle k^2 \rangle} \quad (24)$$

corresponds to the inverse of the largest eigenvalue Λ_M of the adjacency matrix in the annealed network approximation.

For finite networks of size N , all moments of the degree distribution are finite. In this case, we can expand $\Psi(\theta)$ to second order in $p\theta$, corresponding again to the vicinity of the transition point,

$$\theta = 1 - \sum_k \frac{kP(k)}{\langle k \rangle} e^{-kp\theta} \simeq p\theta \frac{\langle k^2 \rangle}{\langle k \rangle} - \frac{\langle k^3 \rangle}{2\langle k \rangle} p^2 \theta^2. \quad (25)$$

Solving the previous equation for nonzero θ , we obtain

$$\theta \simeq \frac{2\langle k \rangle}{p^2 \langle k^3 \rangle} \left[\frac{p}{p_c^{\text{ann}}} - 1 \right] = \frac{2\langle k \rangle}{p^2 \langle k^3 \rangle} [\lambda_0 - 1], \quad (26)$$

where we have defined $\lambda_0 \equiv p/p_c^{\text{ann}}$. This leads to the finite-size expressions

$$b_n = 1 - k_n \frac{2\langle k \rangle}{p \langle k^3 \rangle} [\lambda_0 - 1] \quad (27)$$

and

$$1 - \langle b \rangle \equiv 1 - \frac{\sum_n b_n}{N} = \frac{2\langle k \rangle^2}{p \langle k^3 \rangle} [\lambda_0 - 1]. \quad (28)$$

Notice that the last three equations correspond to expansions close to the critical point, and for this reason they are valid only for $\lambda_0 - 1$ being small, in particular in the regime where $2\langle k \rangle(\lambda_0 - 1)/[p \langle k^3 \rangle]$ is smaller than 1.

2. Average avalanche duration

Close to and below the critical point, i.e., assuming again $p(1 - b_m) \ll 1$, we can write $D_{nm} \simeq pb_n a_{nm}$. Plugging this into Eq. (8), we have

$$T_n = b_n + pb_n \sum_m a_{nm} T_m, \quad (29)$$

which can be written as

$$X_n = 1 + p \sum_m b_m a_{nm} X_m, \quad (30)$$

where $X_n = T_n/b_n$. Given Eq. (30), assuming that X_n is a function of k_n , and in view of Eq. (19), we can make the ansatz $X_n = 1 + \tilde{A}k_n$, with $\tilde{A}$ being a constant to be determined. Inserting the ansatz into Eq. (30), we have

$$\tilde{A}k_n = p \sum_m a_{nm} b_m + \tilde{A}p \sum_m a_{nm} k_m b_m. \quad (31)$$

Using the annealed network approximation, Eq. (19), we have

$$\tilde{A}k_n = pk_n \frac{\langle kb \rangle}{\langle k \rangle} + \tilde{A}pk_n \frac{\langle k^2 b \rangle}{\langle k \rangle}, \quad (32)$$

leading to

$$\tilde{A} = \frac{p \langle kb \rangle}{\langle k \rangle - p \langle k^2 b \rangle}, \quad (33)$$

so that

$$T_n = b_n \left(1 + \frac{p \langle kb \rangle}{\langle k \rangle - p \langle k^2 b \rangle} \right). \quad (34)$$

In the subcritical phase, $b_n = 1$, and we obtain the simple explicit form

$$T_n = 1 + k_n \frac{p \langle k \rangle}{\langle k \rangle - p \langle k^2 \rangle} = 1 + k_n \frac{p}{1 - \lambda_0}. \quad (35)$$

In the supercritical phase, T_n depends on moments of k and b in a more complicated way. Assuming we are close to the critical point, in a finite network, we can use Eq. (27) to obtain

$$\frac{\langle kb \rangle}{\langle k \rangle} = 1 - \frac{2\langle k^2 \rangle}{p \langle k^3 \rangle} (\lambda_0 - 1), \quad (36)$$

$$\frac{\langle k^2 b \rangle}{\langle k \rangle} = \frac{1}{p_c^{\text{ann}}} - \frac{2}{p} (\lambda_0 - 1). \quad (37)$$

Substituting into Eq. (34), we finally have

$$T_n = b_n \left[1 + k_n \left(\frac{p}{\lambda_0 - 1} - \frac{2\langle k^2 \rangle}{\langle k^3 \rangle} \right) \right], \quad (38)$$

which can be written in the fully explicit form

$$T_n = \left[1 - k_n \frac{2(p \langle k^2 \rangle - \langle k \rangle)}{p \langle k^3 \rangle} \right] \times \left[1 + k_n \left(\frac{p \langle k \rangle}{p \langle k^2 \rangle - \langle k \rangle} - \frac{2\langle k^2 \rangle}{\langle k^3 \rangle} \right) \right]. \quad (39)$$

3. Average avalanche size

From Eq. (17), we have $S_n = T_n/b_n$, which translates, from the annealed network solution for T_n , into

$$S_n = 1 + k_n \frac{p\langle kb \rangle}{\langle k \rangle - p\langle k^2 b \rangle}. \quad (40)$$

Hence, in the subcritical phase we have

$$S_n = T_n = 1 + k_n \frac{p}{1 - \lambda_0}, \quad (41)$$

while in the supercritical phase

$$\begin{aligned} S_n &= 1 + k_n \left(\frac{p}{\lambda_0 - 1} - \frac{2\langle k^2 \rangle}{\langle k^3 \rangle} \right) \\ &= 1 + k_n \left(\frac{p\langle k \rangle}{p\langle k^2 \rangle - \langle k \rangle} - \frac{2\langle k^2 \rangle}{\langle k^3 \rangle} \right). \end{aligned} \quad (42)$$

The relation $S_n = T_n/b_n$ implies that in the supercritical case avalanche size and duration are related in a nontrivial way. The explicit expressions point out that the average size (minus 1) is proportional to the degree of the seed close to the epidemic threshold. The same is true for the average duration, but nonlinear effects are stronger in this case.

III. NUMERICAL SIMULATIONS

In this section we perform a systematic comparison between the results of the theoretical approaches presented in the previous section and numerical results obtained by simulating the SIS dynamics on synthetic uncorrelated power-law-distributed networks, built according to the uncorrelated configuration model [23]. The comparison is aimed at validating the accuracy of the theoretical predictions and at identifying the origin of possible inaccuracies in the various approximations performed. The networks considered have a degree distribution $P(k) \sim k^{-\gamma}$ between a minimum value $k_{\min} = 3$ and a maximum $k_{\max} = N^{1/2}$ (for $\gamma < 3$) and $k_{\max} = N^{1/(\gamma-1)}$ for $\gamma > 3$.

We consider infinite all avalanches whose duration is longer than $t_{\inf} = 300$ steps. This threshold, chosen for computational convenience, is larger than the average duration of finite avalanches except for a narrow interval around the critical point. To improve the readability of figures, plotted values of T_n , S_n , and b_n are binned over suitably chosen intervals.

We consider two values of γ corresponding to the two different scenarios occurring for the SIS transition [12]. For $\gamma < 5/2$ the transition is triggered by a subextensive network subset (the max K core), composed of densely mutually interconnected nodes [24]. In this case QMF theory and the annealed network approximation are known to work well, providing an estimate of the position of the epidemic threshold which is asymptotically exact in the large size limit [25,26]. For $\gamma > 5/2$ the epidemic transition is triggered by a different and highly nontrivial mechanism. Nodes with high connectivity (large hubs) considered together with their immediate neighbors form star graphs. When p becomes larger than $p_c^{\text{QMF}} = 1/\Lambda_M \approx 1/\sqrt{k_{\max}}$, the largest of these star graphs becomes able to sustain epidemic activity for a long time, proportional to $\exp(k_{\max})$. For larger values of p other star graphs

FIG. 1. (a) Comparison of $1 - \langle b \rangle$ vs $\lambda = p\Lambda_M$ from numerical simulations for $\gamma = 2.25$ (symbols), with QMF theoretical predictions based on Eq. (1) (dashed lines) for different values of N . (b) and (c) Comparison of b_n from numerical simulations for various $\lambda = p\Lambda_M$ and $N = 10^4$ with annealed network predictions, Eqs. (21) and (27), respectively. In (c), $\lambda_0 = p\langle k^2 \rangle / \langle k \rangle$ corresponds to the annealed network threshold.

can independently sustain activity. QMF theory describes only the behavior of isolated stars. However, the epidemic in each of them survives for a time proportional to $\exp(k_i)$; that is, there is no endemic activity in the system [surviving for a time $\exp(N)$] if stars are isolated [27,28]. As a consequence, the QMF estimate $p_c^{\text{QMF}} = 1/\Lambda_M$ is only a (not strict) lower bound for the actual epidemic threshold. The consideration of long-range interactions among stars, neglected by QMF theory, allows us to understand how endemic activity can be established and to calculate the actual critical threshold $p_c > p_c^{\text{QMF}}$ [16,29].

1. Probability of observing a finite avalanche

In Fig. 1(a) we first compare the probability $1 - \langle b \rangle$ of observing a finite avalanche calculated within the QMF approach [Eq. (1)] with numerical results for $\gamma = 2.25$. It turns out clearly that QMF is very accurate in reproducing the behavior of $1 - \langle b \rangle$. The bending of the numerical curves for small $\lambda - 1$ is due to finite-size effects reflecting the existence of a size-dependent effective threshold. As N increases, the agreement between theory and numerics extends to lower values of $\lambda - 1$, confirming that the QMF threshold estimate is asymptotically exact [25,26]. The other two panels display instead the accuracy of the fully analytical predictions obtained when also the annealed network approximation is implemented. Also in this case we find very good agreement between theory and simulations. Note that no parameter has been fitted.

FIG. 2. (a) Comparison of $1 - \langle b \rangle$ vs $\lambda = p\Lambda_M$ from numerical simulations for $\gamma = 3.50$ (symbols), with QMF theoretical predictions based on Eq. (1) (dashed lines) for different values of N . (b) and (c) Comparison of b_n from numerical simulations for various $\lambda = p\Lambda_M$ and $N = 10^4$ with annealed network predictions, Eqs. (21) and (27), respectively. In (c), $\lambda_0 = p\langle k^2 \rangle / \langle k \rangle$ corresponds to the annealed network threshold.

In Fig. 2 the same analysis is performed for $\gamma = 3.50$. The scenario is completely different. Also in this case the agreement between theory and simulations for what concerns $1 - \langle b \rangle$ holds above a size-dependent threshold, but now the threshold moves toward *larger* values of $\lambda - 1$ as N is increased. This shows that the actual SIS threshold is larger than the QMF prediction, $\lambda_c > \lambda_c^{\text{QMF}} = 1$, and the discrepancy grows with the system size [16,29]. The disagreement is even more evident in Figs. 2(b) and 2(c). In Fig. 2(b), all $b_n = 1$ for $\lambda = 1.1$ despite the QMF prediction that $b_n < 1$ for $\lambda > \lambda_c^{\text{QMF}} = 1$. This is a consequence of the fact that the true value of λ_c is larger than 1.1, so that $\lambda = 1.1$ is actually subcritical instead of supercritical. In Fig. 2(c) we have the additional complication that, when $\lambda = p\Lambda_M$ is 1.1 or 1.2, $\lambda_0 = p\langle k^2 \rangle / \langle k \rangle$ is smaller than 1. This gives rise to the unphysical values larger than 1 along the horizontal axis in Fig. 2(c).

In summary, for $\gamma < 5/2$ both QMF and annealed network theory predictions for the threshold are asymptotically exact, so that one expects theory and simulations to perfectly agree for large N . For $\gamma > 5/2$, instead, both theories predict a threshold value not matching numerics as N grows. In that case the predictions for b_n and $\langle b \rangle$ work only sufficiently inside the supercritical phase.

2. Average avalanche duration and size

In Fig. 3 we report a comparison between QMF predictions (along the horizontal axis) and numerical results (vertical axis) for $\gamma = 2.25$. Results are presented for both the average

FIG. 3. Comparison of numerical simulations with QMF theoretical predictions [Eqs. (9), and (16)] for $\gamma = 2.25$ and $N = 10^4$. (a) T_n subcritical, (b) S_n subcritical, (c) T_n supercritical, and (d) S_n supercritical.

duration T_n (left column) and the average avalanche size S_n (right column). In the subcritical case excellent agreement is found for S_n for any value of λ , while the same is true only for durations up to $T_n \approx 1$, corresponding to small values of λ . For larger durations a disagreement starts to appear for T_n . This difference can be rationalized based on the different treatments of Sec. II A. The prediction for S_n is valid for any size, without any small-size assumptions. This failure of QMF theory in predicting the avalanche duration can be, instead, attributed to the breakdown of the assumption made to obtain Eq. (7). Close to the transition point, critical slowing down leads to a slow decay of $f_n(t)$ toward zero, and it is not appropriate to assume linearization to hold starting from $t = 1$.

For the average avalanche size in the supercritical case [Fig. 3(d)], large discrepancies between theory and simulations occur for λ slightly larger than 1, while excellent agreement is found again for larger λ . This finding can be ascribed to the intrinsic impossibility of discriminating precisely finite or infinite avalanches in networks of finite size. At the critical point the distribution of avalanche times features a power-law decay plus incipient infinite avalanches, characterized by long, but still finite, durations. Some of them are classified as finite according to the criterion $T < t_{\text{inf}}$ and therefore contribute to T_n increasing its value by a large amount. Reducing t_{inf} alleviates this problem, but at the price of risking to spuriously consider as infinite avalanches those belonging to the power-law decay. This difficulty is naturally mitigated by considering larger systems or going to larger λ as the separation between the two components of the duration distribution becomes more pronounced. For the average avalanche duration in the supercritical case [Fig. 3(c)] this intrinsic problem with finite size (which tends to increase T_n) competes with the breakdown of the linearization for f_n (which tends to decrease T_n). As a consequence, numerical results are not far from QMF predictions. Figure 4 reports the same analysis for $\gamma = 3.50$; the same qualitative scenario

FIG. 4. Comparison of numerical simulations with theoretical predictions [Eqs. (9) and (16)] for $\gamma = 3.50$ with $N = 10^4$. (a) T_n subcritical, (b) S_n subcritical, (c) T_n supercritical, and (d) S_n supercritical.

emerges concerning the agreement between QMF theory and simulations. This is somewhat surprising, given the known failure of QMF in describing accurately the transition in this case. However, this failure becomes most evident for very large networks, while only moderate values of N are considered here [26].

Further insight into what happens for long subcritical avalanches close to the threshold is provided in Fig. 5, where we plot the average avalanche size versus the corresponding duration, obtained from numerical simulations. As we can see here, for small values of λ , the QMF prediction $S_n \sim T_n$ is well

FIG. 5. Plot of the avalanche size S_n as a function of the avalanche duration T_n for (a) $\gamma = 2.25$ and (b) $\gamma = 3.50$ in the subcritical regime. (c) and (d) The avalanche size S_n (normalized by its maximum value) as a function of the square of the avalanche duration T_n normalized by its maximum value for $\gamma = 2.25$ and $\gamma = 3.50$, respectively. Data are from numerical simulations in networks of size $N = 10^4$.

FIG. 6. Comparison of numerical simulations with annealed network predictions for $\gamma = 2.25$ and $N = 10^4$. (a) T_n subcritical [Eq. (35)], (b) S_n subcritical [Eq. (41)], (c) T_n supercritical [Eq. (39)], and (d) S_n supercritical [Eq. (42)].

fulfilled. Closer to the critical point, on the other hand, we observe a departure from the linear behavior, which reflects the mismatch between the poor agreement with the QMF theory for large λ of the avalanche duration T_n and the very good agreement of the avalanche size S_n . Numerically, we observe instead a good fit for λ close to 1 to the form $S_n \sim T_n^2$ [see Figs. 5(c) and 5(d)], which is a consequence of the critical scalings of $p_n(s)$ and $p_n(t)$ [15].

Figures 6 and 7 analyze instead the agreement between simulations and the theoretical results obtained using the annealed network approximation for both T_n [Eqs. (35) and (39)] and S_n [Eqs. (41) and (42)]. For $\gamma = 2.25$ we find very good agreement throughout the whole phase diagram, except for a critical region around the epidemic threshold. For $\gamma = 3.50$,

FIG. 7. Comparison of numerical simulations with annealed network predictions for $\gamma = 3.50$ and $N = 10^4$. (a) T_n subcritical [Eq. (35)], (b) S_n subcritical [Eq. (41)], (c) T_n supercritical [Eq. (39)], and (d) S_n supercritical [Eq. (42)].

FIG. 8. Dependence of the avalanche average duration and size as a function of b_n in the supercritical regime. (a) T_n and (b) S_n for $\gamma = 2.25$. (c) T_n and (d) S_n for $\gamma = 3.50$.

the fit is very good in the subcritical regime and rather poor away from the critical point in the supercritical regime. This is due to the fact that the annealed network threshold p_c^{ann} largely overestimates the actual threshold. A consequence of this overestimate is that for λ slightly larger than 1, Eqs. (39) and (42) predict negative values of T_n and S_n . A better agreement between theory and numerical results is expected for larger values of λ , which will be less affected by the mismatch between the prediction of the threshold and its actual value.

Finally, in Fig. 8 we report the dependence of the average avalanche duration T_n and size S_n in the supercritical regime as a function of the probability of observing a finite avalanche b_n . For both $\gamma = 2.25$ and $\gamma = 3.50$, we find continuously decreasing functions of b_n , with large but not huge fluctuations. The monotonicity allows us to assume, as a rule of thumb, that if a node has a larger probability $1 - b_n$ to originate an infinite avalanche than a second node, the first will also give rise, on average, to longer and larger *finite* avalanches. In other words, ranking nodes according to b_n , T_n , or S_n produces the same ordering, apart from fluctuations.

IV. CONCLUSIONS

In this paper we presented a thorough investigation, both analytical and numerical, of the problem of identifying influential spreaders for SIS dynamics on generic networks. We first applied to this problem the theoretical QMF framework introduced by Larremore *et al.* [15]. In this way we were able to write down closed equations whose numerical solution allows us to calculate all the observables of interest in the whole phase diagram. Within this approach a further step, consisting of the implementation of the annealed network approximation, allows us to derive explicit analytical predictions valid below and above (but close to) the critical point. These predictions point out the importance of degree centrality in determining the spreading influence. A comparison of these results with analytical simulations, performed on synthetic networks with power-law-distributed degrees $P(k) \sim k^{-\gamma}$, confirms the

substantial accuracy of the theoretical approaches below the threshold, in particular for what concerns the avalanche size. Above the expected QMF threshold discrepancies arise, which tend to disappear inside the active region of the phase diagram.

Let us discuss in detail strengths and limitations of the theoretical approaches considered here. The QMF theory of Larremore *et al.* [15] is based on three approximations: the neglect of dynamical correlations among neighbors' states, the locally treelike assumption that allows us to write down Eqs. (1) and (2), and the linearization entailed in Eq. (3) for the derivation of T_n . The annealed network approximation adds another assumption on top of them. For strongly heterogeneous unclustered networks, such as those generated by the uncorrelated configuration model with $\gamma < 5/2$, the first two assumptions are essentially correct. Studies on SIS dynamics have shown that even the annealed network approximation works very well in this case. The linearization holds instead only if one is sufficiently far from the epidemic threshold. Close to it, the regime with exponential decay into the stationary state is preceded by a long transient during which the decay is power law (critical slowing down), f_n is not small, and linearization does not hold; the dominating contribution to the overall duration of avalanches comes from this long preasymptotic regime. For this reason predictions for T_n are not consistent with simulations for values of p close to the threshold. When $\gamma > 5/2$, the agreement between our theoretical results and simulations is expected to be poorer. The reason is that, already, the first approximation, the neglect of dynamical correlations, fails to capture the complex physical mechanisms underlying the epidemic transition and thus introduces large errors. See Ref. [12] for more on this. A fundamental consequence is that, already, the QMF estimate of the epidemic threshold is qualitatively and quantitatively incorrect. This leads to the expectation that predictions for influential spreaders in the supercritical case are largely off target. The relatively small errors observed here are believed to be an effect of the relatively small network sizes considered. The additional annealed network approximation, which does not work for $\gamma > 5/2$, makes the theoretical approaches perform even worse. Recent results [29] have pointed out the great complexity of the physical mechanisms underlying the epidemic transition in random networks with $\gamma > 5/2$. This understanding constitutes the necessary groundwork for reaching a satisfactory predictive ability for influential spreaders also in this case, a goal that remains to be reached. Similarly, the extension of this approach to networks exhibiting more complicated features, such as clustering, correlations, and mesoscopic structures, remains an interesting open issue.

Finally, it is worth remarking that our results have been derived for a particular discrete-time version of SIS dynamics, rather different from the usual continuous-time version. The present version, with parallel dynamics and recovery time fixed to 1, allows a freshly infected node to immediately reinfect the node that infected it in the preceding time step. This kind of process is, instead, strongly hindered in continuous-time SIS, where the neighbor j that infected node i remains infected for a while, so that the number of connections of i available for further spreading is effectively $k_i - 1$. This kind of dynamical correlation has strong effects for influential spreaders in SIR dynamics. In that case degree centrality is a

good approximation of the spreading influence of individual nodes, but close to the transition a better approximation is provided by the nonbacktracking centrality [8,11,30], which may differ markedly from the former. The quest for accurate predictors of spreading influence for continuous-time SIS dynamics close to the transition is an interesting open avenue for further research.

ACKNOWLEDGMENTS

We acknowledge financial support from the Spanish government's MINECO, under Project No. FIS2016-76830-C2-1-P. R.P.-S. acknowledges additional financial support from ICREA Academia, funded by the Generalitat de Catalunya regional authorities.

-
- [1] *Complex Spreading Phenomena in Social Systems*, edited by S. Lehmann and Y.-Y. Ahn (Springer, Cham, 2018).
 - [2] S. Pei, L. Muchnik, J. S. Andrade, Jr., Z. Zheng, and H. A. Makse, *Sci. Rep.* **4**, 5547 (2014).
 - [3] L. Lu, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, *Phys. Rep.* **650**, 1 (2016).
 - [4] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, *Nat. Phys.* **6**, 888 (2010).
 - [5] S. B. Seidman, *Soc. Networks* **5**, 269 (1983).
 - [6] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, *Phys. A* **391**, 1777 (2012).
 - [7] Q. Li, T. Zhou, L. Lü, and D. Chen, *Phys. A* **404**, 47 (2014).
 - [8] F. Radicchi and C. Castellano, *Phys. Rev. E* **93**, 062314 (2016).
 - [9] K.-i. Hashimoto, *Automorphic Forms and Geometry of Arithmetic Varieties* (Mathematical Society of Japan, Tokyo, 1989), pp. 211–280.
 - [10] F. Krzakala and C. Moore, *Proc. Natl. Acad. Sci. USA* **110**, 20935 (2013).
 - [11] B. Min, *Eur. Phys. J. B* **91**, 18 (2018).
 - [12] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
 - [13] Q. Liu and P. Van Mieghem, in *Complex Networks and Their Applications V*, edited by H. Cherifi, S. Gaito, W. Quattrociocchi, and A. Sala (Springer, Cham, 2017), pp. 511–521.
 - [14] P. Holme and L. Tupikina, *New J. Phys.* **20**, 113042 (2018).
 - [15] D. B. Larremore, M. Y. Carpenter, E. Ott, and J. G. Restrepo, *Phys. Rev. E* **85**, 066131 (2012).
 - [16] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. Lett.* **111**, 068701 (2013).
 - [17] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, in *22nd International Symposium on Reliable Distributed Systems (SRDS'03)* (IEEE Computer Society, Los Alamitos, CA, 2003), pp. 25–34.
 - [18] P. Van Mieghem, J. Omic, and R. E. Kooij, *IEEE/ACM Trans. Networking* **17**, 1 (2009).
 - [19] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Europhys. Lett.* **89**, 38009 (2010).
 - [20] C. Castellano and R. Pastor-Satorras, *Phys. Rev. Lett.* **105**, 218701 (2010).
 - [21] J. Goldenberg, B. Libai, and E. Muller, *Mark. Lett.* **12**, 211 (2001).
 - [22] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
 - [23] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, *Phys. Rev. E* **71**, 027103 (2005).
 - [24] C. Castellano and R. Pastor-Satorras, *Sci. Rep.* **2**, 00371 (2012).
 - [25] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. E* **86**, 041125 (2012).
 - [26] D. H. Silva, S. C. Ferreira, W. Cota, R. Pastor-Satorras, and C. Castellano, *Phys. Rev. Res.* **1**, 033024 (2019).
 - [27] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, *Phys. Rev. Lett.* **109**, 128702 (2012).
 - [28] H. K. Lee, P.-S. Shim, and J. D. Noh, *Phys. Rev. E* **87**, 062812 (2013).
 - [29] C. Castellano and R. Pastor-Satorras, *Phys. Rev. X* **10**, 011070 (2020).
 - [30] T. Martin, X. Zhang, and M. E. J. Newman, *Phys. Rev. E* **90**, 052808 (2014).