




High prevalence regimes in the pair-quenched mean-field theory for the susceptible-infected-susceptible model on networks

Diogo H. Silva ^{1,*} Francisco A. Rodrigues ^{2,†} and Silvio C. Ferreira ^{1,3,‡}

¹*Departamento de Física, Universidade Federal de Viçosa, 36570-900 Viçosa, Minas Gerais, Brazil*

²*Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 13566-590 São Carlos, São Paulo, Brazil*

³*National Institute of Science and Technology for Complex Systems, 22290-180 Rio de Janeiro, Rio de Janeiro, Brazil*



(Received 3 June 2020; accepted 30 June 2020; published 23 July 2020)

Reckoning of pairwise dynamical correlations significantly improves the accuracy of mean-field theories and plays an important role in the investigation of dynamical processes in complex networks. In this work, we perform a nonperturbative numerical analysis of the quenched mean-field theory (QMF) and the inclusion of dynamical correlations by means of the pair quenched mean-field (PQMF) theory for the susceptible-infected-susceptible model on synthetic and real networks. We show that the PQMF considerably outperforms the standard QMF theory on synthetic networks of distinct levels of heterogeneity and degree correlations, providing extremely accurate predictions when the system is not too close to the epidemic threshold, while the QMF theory deviates substantially from simulations for networks with a degree exponent $\gamma > 2.5$. The scenario for real networks is more complicated, still with PQMF significantly outperforming the QMF theory. However, despite its high accuracy for most investigated networks, in a few cases PQMF deviations from simulations are not negligible. We found correlations between accuracy and average shortest path, while other basic network metrics seem to be uncorrelated with the theory accuracy. Our results show the viability of the PQMF theory to investigate the high-prevalence regimes of recurrent-state epidemic processes in networks, a regime of high applicability.

DOI: [10.1103/PhysRevE.102.012313](https://doi.org/10.1103/PhysRevE.102.012313)

I. INTRODUCTION

Pairwise approximation constitutes a valuable tool recurrently used for understanding dynamical processes in graphs (networks or lattices) and, particularly, epidemic spreading on the top of complex networks [1–3]. This approach outperforms ordinary mean-field approximations, extending dynamical equations from one site to the pair level [4]. Extensions to higher-order methods using n -cluster approximations [5] can lead to more accurate theories at the cost of increasing the theoretical complexity. While being of limited application for low-dimensional systems near critical phase transitions [4], pair approximations can be remarkable improvements with respect to the one-site theory either if we are not too close to the transition [6] or if the system dimension is large such as the case of random graphs [7,8].

For dynamical processes on the top of complex networks, heterogeneities play a central role [9,10] that has to be taken into account to reproduce the most fundamental results [11,12]. Particularly, the interplay between structural heterogeneity and dynamical correlations has been investigated using heterogeneous pair approximations [7,8,13–15]. We consider the susceptible-infected-susceptible (SIS) model [9], whose dependence on heterogeneities serves as a reference for many other dynamical processes [9,16]. In the SIS model individuals are represented by vertices of a network and can be

in either susceptible or infected states. Infected vertices heal spontaneously at rate μ and infect their susceptible neighbors at rate λ per contact. A central aspect of spreading phenomena is the epidemic threshold λ_c , above which an extensive fraction of the population is infected or, in other words, the epidemic prevalence is finite.

Heterogeneities of networks can change the behavior of the threshold drastically. If the network possesses a heavy-tailed degree distribution in the form of a power law $P(k) \sim k^{-\gamma}$, the epidemic threshold of the SIS model is 0 in the thermodynamical limit when the network size $N \rightarrow \infty$ [17–19]. This involves a very special type of transition from an active and fluctuating to an absorbing state [20,21], which can be knocked out with small modifications of the SIS dynamics [22,23]. The SIS transition on uncorrelated power-law networks can be of two types [20–22]: If the degree exponent is small ($\gamma < 2.5$) the activation is triggered by a densely connected core identified by the maximal index of a k -core decomposition [20]. If the degree exponent is large ($\gamma > 2.5$), then the activation is ruled by the hubs. The latter involves long-term epidemic activity on star subgraphs, composed of a single hub (the center) and its k_{hub} nearest neighbors (the leaves), through a feedback mechanism where the hub infects the leaves, which in turn reinfect the hub. This activity must last for sufficiently long times to permit the mutual activation of hubs which are not directly connected (long-range) [17,18,21].

Heterogeneities can be included in mean-field approximations in different forms [10]. Two widely used approximations are the heterogeneous mean-field (HMF) [11] and quenched mean-field (QMF) [12,24] theories. The former consists of a

*diogo.henrique@ufv.br

†francisco@icmc.usp.br

‡silviojr@ufv.br

coarse-graining where only the degree of the nodes and the statistical degree correlations are included in the dynamical equations for the probability that a node is infected [11,25] and neglects the dynamical correlations. The latter includes the full connectivity structure of the networks but still neglects dynamical correlations assuming that the states of nearest neighbors are independent. Due to the aforementioned nature of the SIS activation mechanisms, the explicit inclusion of the network connectivity structure, as in the QMF approach, is imperative to construct mean-field theories of the SIS model since heterogeneous mixing, which corresponds to an annealed network [20], will destroy some localization effects such as, for example, the self-sustained activity in a star subgraph.

Despite the detailed microscopic description of the QMF theory, neglecting dynamical correlations in the SIS model can lead to modest accuracy with significant deviations from simulations [26–28] if the epidemic involves, for example, activation localized in the hubs that spreads to the rest of network [17,20]. Indeed, the threshold predicted by the QMF theory for the SIS model, given by the inverse of the largest eigenvalue of the adjacency matrix [12,19] (see Sec. II for details), involves a localized phase on random power-law networks with degree exponent $\gamma > 2.5$ [29]. Dynamical correlations reckoned by individual pairwise interactions greatly improve the predictions of the epidemic thresholds of the QMF approach in the so-called pair QMF (PQMF) theory [27]; see Sec. II for details. Indeed, PQMF theory [2,3,28,30] and modified versions of it [31–34] have been intensively investigated recently. The asymptotic scaling exponent of the threshold as a function of the network size is unchanged when the pair approximation is included in the QMF theory [27,28].

Since PQMF theory has been mainly analyzed perturbatively in the limit of very low prevalence to investigate the position of the epidemic thresholds, Matamalas *et al.* [31] claimed that it has limitations to compute high-epidemic-incidence regimes and proposed that a microscopic Markov chain approach (MMCA) [35], which is a discrete-time version of the QMF theory, could be used instead. However, a nonperturbative approach is possible through numerical integration of both QMF and PQMF dynamical equations. Since large discrepancies between discrete- and continuous-time approaches can be present in the SIS dynamics [36], a nonperturbative analysis of QMF and PQMF theories is necessary. We develop nonperturbative analyses of both QMF and PQMF theories for SIS on networks using numerical integration of the corresponding dynamical equations. We consider both large synthetic networks generated with the Weber-Porto model [37] and real networks with different levels of degree correlation. We address regimes not asymptotically close to the epidemic threshold since the mean-field theories fail in predicting very low densities of infected vertices [28,38]. However, numerical analyses support that this asymptotic scaling is confined to a small interval near $\lambda = \lambda_c \rightarrow 0$ such that the mean-field theories are still applicable beyond this region (see Fig. 1). We report that PQMF theory predicts very accurately the epidemic prevalence in synthetic networks for all ranges of degree exponent ($\gamma < 2.5$ and $\gamma > 2.5$) and correlations (assortative, disassortative, and uncorrelated) investigated, while QMF theory presents significant deviations

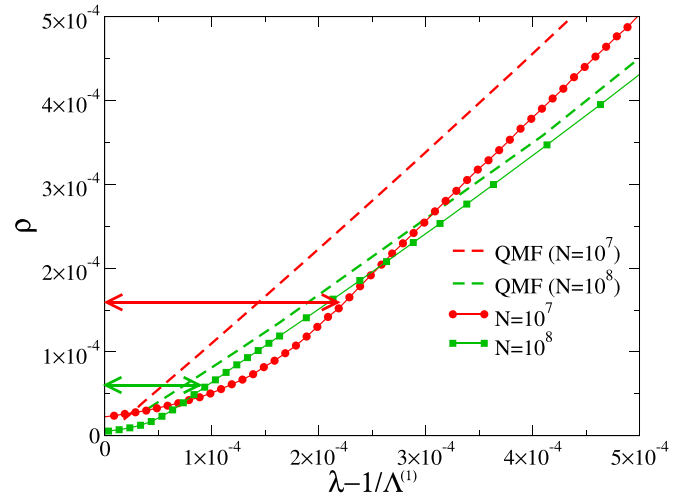


FIG. 1. Epidemic prevalence around the transition threshold $\lambda_c = 1/\Lambda^{(1)}$ for uncorrelated configuration model networks with $\gamma = 2.3$, $k_{\min} = 3$, and $k_{\max} = 2\sqrt{N}$. Simulations are represented by solid lines with symbols, and numerical integrations of the QMF equation are shown by the dashed lines. Horizontal arrows indicate the interval where curves depart from linearity.

for $\gamma > 2.5$. For real networks, in general, PQMF theory considerably outperforms QMF but also presents nonnegligible deviations from simulations in some cases.

The remainder of the paper is organized as follows. The mean-field theories used in this work are discussed in Sec. II. Comparison of numerical integration and stochastic simulations is performed in Sec. III, while our concluding remarks are presented in Sec. IV. Appendixes A, B, and C present technical details of our numerical analyses.

II. MEAN-FIELD THEORIES

We investigate the SIS model in a connected, undirected, and unweighted network with $i = 1, \dots, N$ vertices whose structure is encoded in the adjacency matrix A_{ij} defined by $A_{ij} = 1$ if i and j are connected and $A_{ij} = 0$ otherwise. The healing rate is fixed to $\mu = 1$ without loss of generality.

The probability that a vertex i is infected, represented by ρ_i , evolves as [27]

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda \sum_j \phi_{ij} A_{ij}, \quad (1)$$

where ϕ_{ij} is the probability that a vertex i is susceptible and its nearest neighbor j is infected. Equation (1) is exact but not closed. A closed system is obtained taking the one-site approximation $\phi_{ij} \approx \rho_i(1 - \rho_j)$ that corresponds to the QMF theory [24,29]:

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda(1 - \rho_i) \sum_{j=1}^N A_{ij} \rho_j. \quad (2)$$

The QMF epidemic threshold is given by $\lambda_c^{\text{QMF}} \Lambda^{(1)} = 1$, where $\Lambda^{(1)}$ is the largest eigenvalue of the adjacency matrix A_{ij} .

The PQMF theory includes dynamical correlations considering the evolution of ϕ_{ij} which, in its complete but not closed form, depends on the triplets; see Ref. [27]. The PQMF theory consists of approximating the triplets $[A_i, B_j, C_l]$ in which i and l are both connected to j by

$$[A_i, B_j, C_l] \approx \frac{[A_i, B_j][B_j, C_l]}{[B_j]}. \quad (3)$$

Here A , B , and C are the states of the vertices that, in the SIS case, can be either infected or susceptible. The approximation given by Eq. (3) considers that (i, j, l) does not form a triangle, i.e., i and l are connected to j but not to each other. Actually, the effects of clustering have been recently investigated [34] for connected random networks and it was shown that even in networks with a high cluster coefficient and plenty of triangles, the approximation given by Eq. (3) performs very well for the steady state. The final PQMF equation for ϕ_{ij} becomes

$$\begin{aligned} \frac{d\phi_{ij}}{dt} = & -(2 + \lambda)\phi_{ij} + \rho_j + \lambda \sum_l \frac{\omega_{ij}\phi_{jl}}{1 - \rho_j} (A_{jl} - \delta_{il}) \\ & - \lambda \sum_l \frac{\phi_{ij}\phi_{il}}{1 - \rho_i} (A_{il} - \delta_{lj}), \end{aligned} \quad (4)$$

in which $\omega_{ij} = 1 - \phi_{ij} - \rho_i$. Equations (1) and (4) form a closed system of $N + M$ equations where $M = \frac{1}{2} \sum_j k_j$ is the number of edges of the network. More details of the derivation are given in Ref. [27].

The epidemic threshold within the PQMF framework is given by the transcendent equation $\lambda_c \Omega^{(1)}(\lambda_c) = 1$, [28] where $\Omega^{(1)}$ is the largest eigenvalue of the weighted adjacency matrix B_{ij} given by

$$B_{ij} = \frac{2 + \lambda}{2\lambda + 2} \frac{A_{ij}}{1 + \frac{\lambda^2 k_i}{2\lambda + 2}}. \quad (5)$$

See [28] for the derivation of Eq. (5).

Very close to the threshold, the epidemic prevalence ρ , defined as $\rho = \frac{1}{N} \sum_i \rho_i$, approaches 0 following a power law in the form $\rho \simeq a_1 (\lambda - \lambda_c)^\beta$, where β is a critical exponent [4] and a_1 is prefactor that may depend on the network size. In both QMF [29,39] and PQMF [28] theories it can be shown that $\beta_{\text{QMF}} = \beta_{\text{PQMF}} = 1$, while a_{QMF} or a_{PQMF} can be expressed in terms of the principal eigenvector (PEV) $\{v_i^{(1)}\}$ of either A_{ij} or $B_{ij}(\lambda_c)$, respectively, as [28,29]

$$a = \frac{\sum_{i=1}^N v_i^{(1)}}{N \sum_{i=1}^N [v_i^{(1)}]^3}. \quad (6)$$

These results are straightforwardly derived when the network presents a spectral gap $\Lambda^{(1)} \gg \Lambda^{(2)}$, where $\Lambda^{(2)}$ is the second largest eigenvalue of the adjacency matrix (see, e.g., [28] and [29]). However, it was shown that $\beta_{\text{QMF}} = 1$ is always true [40] and the same is expected for the PQMF theory since pair approximations should not change the universality class predicted by the one-vertex theory [5].

The mean-field exponent $\beta = 1$ does not match the rigorous results obtained by Mountford *et al.* [38] in the thermody-

namical limit $N \rightarrow \infty$, where

$$\rho \sim \begin{cases} \lambda^{\frac{1}{3-\gamma}} & \text{if } 2 < \gamma < 5/2, \\ \frac{\lambda^{2\gamma-3}}{(\ln \frac{1}{\lambda})^{\gamma-2}} & \text{if } 5/2 < \gamma < 3, \\ \frac{\lambda^{2\gamma-3}}{(\ln \frac{1}{\lambda})^{2\gamma-4}} & \text{if } \gamma > 3, \end{cases} \quad (7)$$

according to which $\beta > 1$ for any $\gamma > 2$. For large networks with $\gamma < 5/2$, where the epidemic threshold is very accurately reproduced by the QMF theory, the numerical integration performed in Ref. [28] confirms the deviation from the exact result for λ approaching $\lambda_c^{\text{QMF}} = \frac{1}{\Lambda^{(1)}}$, while stochastic simulations are in agreement with the rigorous results. However, the simulations show a preasymptotic behavior fully consistent with $\beta_{\text{QMF}} = 1$. On a linear scale [41], the region that departs from linearity is squeezed around $\lambda = \lambda_c \rightarrow 0$ as the network size increases, as indicated by the horizontal arrows in Fig. 1, in which simulations and QMF theory are compared. The slope of the linear region decreases with size since a_{QMF} also does: We found $a_{\text{QMF}} = 0.00382$ for $N = 10^7$ and $a_{\text{QMF}} = 0.00130$ for $N = 10^8$. Finally, we see that QMF is not able to capture quantitatively the amplitude of the linear region observed in simulations reinforcing the need of nonperturbative analyses of the PQMF theory.

A closed solution for Eqs. (1) and (4) can be derived for the particular case of homogeneous networks where $P(k) = \delta_{k,m}$ for which $\rho_i = \rho$ and $\phi_{ij} = \phi$. The expression for stationary epidemic prevalence is [42]

$$\bar{\rho} = \frac{\lambda - \lambda_c}{m^{-1} + \lambda - \lambda_c}, \quad \lambda_c = \frac{1}{m-1}. \quad (8)$$

III. RESULTS

We numerically integrated QMF and PQMF equations using a fourth-order Runge-Kutta method with time step $\delta t = 10^{-2}$ to 10^{-1} . Initial conditions consistent with the exact closure equations relating pairwise and single vertex probabilities such as $[S_i, I_j] + [I_i, I_j] = [I_j]$ must be chosen and the steady state is insensitive to a particular choice. We performed stochastic simulations of the SIS dynamics on networks using an optimized Gillespie algorithm [44]; see Appendix C. The absorbing states, which are rigorously the unique real stationary state in finite-size networks, were circumvented using quasistationary simulations [45]; see Appendix B.

A. Synthetic networks

A simple metrics to quantify the correlations is the average degree of the nearest neighbors of vertices with a given degree k [9,46], represented by $\kappa_{\text{nn}}(k)$. The functional form of $\kappa_{\text{nn}}(k)$ reveals correlation patterns of the network. If κ_{nn} is an increasing function of k , the network presents assortative correlations where vertices of similar degree tend to be connected. Conversely, if κ_{nn} decreases with k the network has disassortative correlations where vertices of high degree tend to be connected with vertices of low degree. Finally, if κ_{nn} does not depend on the degree, the network is said to be uncorrelated or neutral and assumes the value $\kappa_{\text{nn}} = \langle k^2 \rangle / \langle k \rangle$ [9].

We investigated networks with distributions $P(k) \propto k^{-\gamma}$ and $k = k_{\text{min}}, \dots, k_{\text{max}}$, where $k_{\text{min}} = 3$. For $\gamma < 3$ we used

$k_{\max} = 2\sqrt{N}$, which permits us to build networks without degree correlation with the uncorrelated configuration model [47]. The factor 2 helps to accelerate the convergence to the asymptotic limit where both $N \rightarrow \infty$ and $k_{\max} \rightarrow \infty$ [28]. For $\gamma > 3$, a rigid cutoff given by $P(k_{\max})N = 1$ [48] was used to suppress multiple (localized) transitions [49] and facilitate threshold determination. Degree correlations were included using the benchmark model proposed by Weber and Porto [37], hereafter called the Weber-Porto configuration model; see Appendix A. The dependence $\kappa_{nn}(k) \propto k^\alpha$ was investigated where $\alpha < 0$, $\alpha = 0$, and $\alpha > 0$ correspond to disassortative, neutral, and assortative correlations, respectively. All investigated networks are connected.

Epidemic prevalences obtained from theories and stochastic simulations are compared in Fig. 2. PQMF outperforms QMF theory in all investigated cases, as intuitively expected since the former includes higher-order correlations. While QMF theory deviates from simulation for regimes of high densities of infected vertices, PQMF cannot be distinguished from simulations in the presented scales. Under low-density regimes, QMF and PQMF agree very well and are indistinguishable from simulations for $\gamma = 2.3$ [Fig. 2(a)], while larger deviations of QMF can be seen for larger values of γ [Figs. 2(b) and 2(c)]. The accuracy of the theories at low prevalence is better for assortative and worse for disassortative networks compared with the neutral case. Another interesting dependence on the assortativity can be observed in these curves. For low prevalences, assortative and disassortative networks possess, respectively, higher and lower densities compared with the uncorrelated networks. At high prevalences, the converse is observed, where disassortative networks present higher densities than the assortative and neutral networks. The same behavior is observed for all values of γ , indicating that it is related to the degree correlations. The behavior at low densities can be explained in terms of the reduced capacity to transmit infection when hubs are surrounded by low-degree vertices in the disassortative case rather than being directly connected with higher probability in the assortative case. We cannot provide simple arguments for the inverted dependence on the assortativity degree at high densities, but it is very precisely reproduced by the PQMF theory. We also simulated the SIS on larger networks with $N = 10^7$ vertices and the level of accuracy of the mean-field theories is similar.

We also investigate the role of the heterogeneity comparing the PQMF theory with a pair homogeneous mean-field (PHMF) theory using Eq. (8) with m replaced by the average degree $\langle k \rangle$ of the network [50,51]. The densities of infected vertices obtained in both pairwise approaches are shown in Fig. 3. Beyond the expected discrepancy for describing the low-prevalence regimes, since one theory predicts a finite and the other a vanishing threshold, the regime of high epidemic prevalence is affected by the inclusion of heterogeneity. As one would expect, the more heterogeneous networks present the larger discrepancies between homogeneous and heterogeneous theories.

B. Real networks

Real networks usually present some degree of correlation, and in many cases, the patterns can be quite complex,

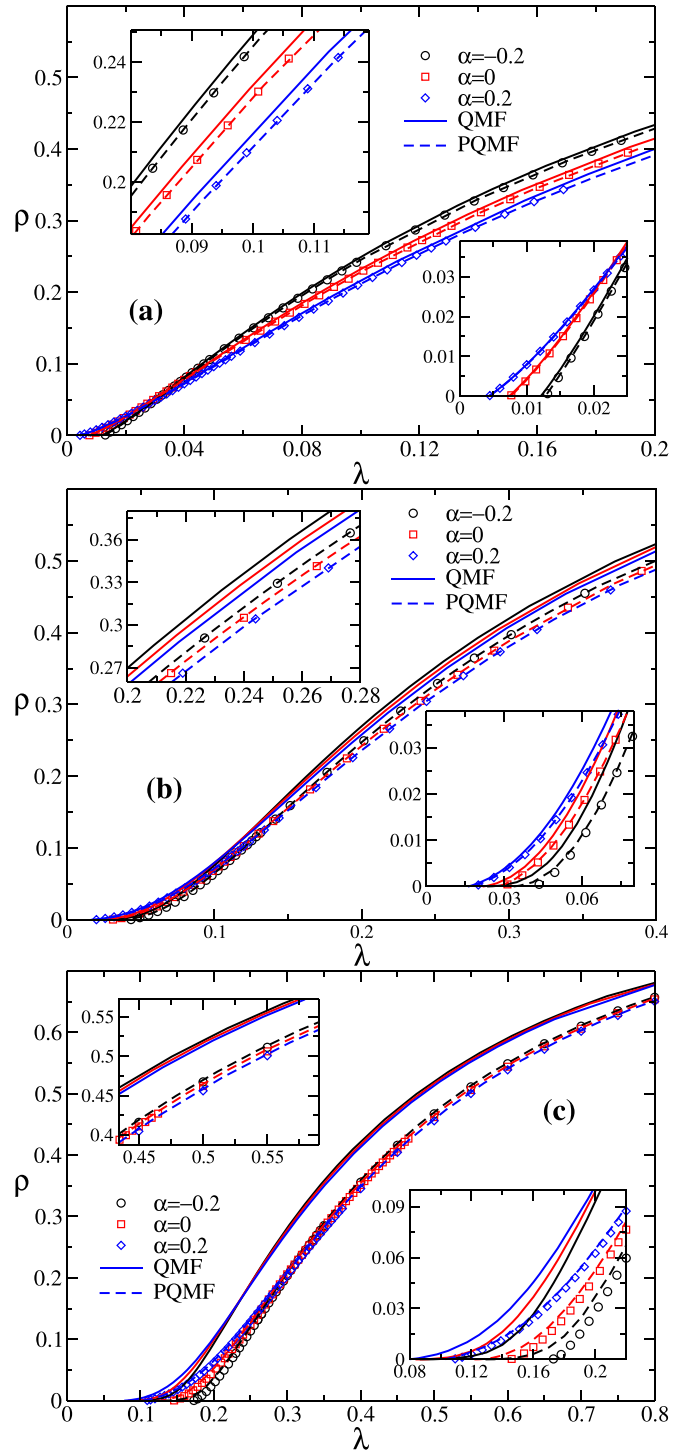


FIG. 2. Epidemic prevalence as a function of the infection rate for Weber-Porto configuration model networks with $N = 10^6$ vertices, degree exponent (a) $\gamma = 2.3$, (b) $\gamma = 2.8$, and (c) $\gamma = 3.5$, and different levels of degree correlations. Symbols represent stochastic simulation; solid and dashed lines, the numerical integration of the QMF and PQMF theories, respectively. Bottom inset: Zoom-in of low prevalence. Top inset: Zoom-in of high prevalence.

exhibiting both assortative and disassortative correlations for distinct ranges of degree [9,52]. Therefore the comparison between mean-field theories and simulations are necessary in order to determine to which extent the accuracy observed in

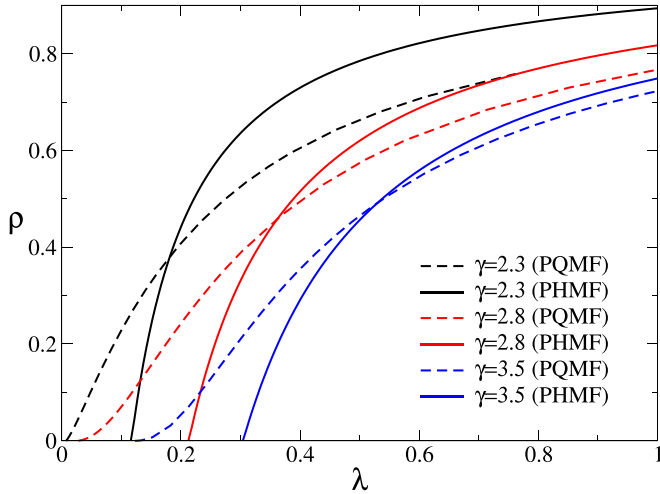


FIG. 3. Comparison of PQMF (dashed lines) and PHMF (solid lines) theories for uncorrelated networks with different levels of heterogeneity. The network size is $N = 10^6$.

synthetic networks holds in the real-world counterparts. We selected some networks with different levels of heterogeneity, sizes, and correlations recently used in the investigation of epidemic processes [28,53,54]. Only the largest connected components are used if there is more than one. For detailed information about the original references for all the networks see Refs. [55,56].

Figure 4 presents the prevalence as a function of the infection rate for 12 real networks. We remark that data asymptotically close to the epidemic threshold are known to mismatch simulations [28] and are beyond the scope of the present work. In some cases, QMF and PQMF are indistinguishable from each other and agree almost perfectly with simulations in the scale presented in these figures. In other cases, QMF theory deviates considerably from simulations while PQMF remains accurate. In order to quantify the differences we define the relative deviation of densities obtained in simulations (ρ_{sim}) with the QMF theory (ρ_{QMF}) as

$$\eta_{QMF} = \frac{\int_{\lambda_1}^{\lambda_2} [\rho_{QMF}(\lambda) - \rho_{sim}(\lambda)] d\lambda}{\int_{\lambda_1}^{\lambda_2} \rho_{sim}(\lambda) d\lambda}, \quad (9)$$

where λ_1 and λ_2 are the initial and final infection rates in the simulations presented in Fig. 4, and an equivalent definition of η_{PQMF} for the PQMF theory. The intervals $[\lambda_1, \lambda_2]$ were chosen to include sufficiently high densities to guarantee that the effects of the absorbing states are small and prevalences are not close to 1. The relative deviations are listed in Table I. The positivity of η shows that the mean-field theories overestimate the density obtained in simulations as expected since dynamical correlations, pruned in mean-field theories, reduce the spreading capacity of the epidemic process. As can be seen, we have $\eta_{PQMF} \ll \eta_{QMF}$ such that PQMF is always more precise than QMF. Three networks present significant deviations of the PQMF theory from the simulations, namely, for

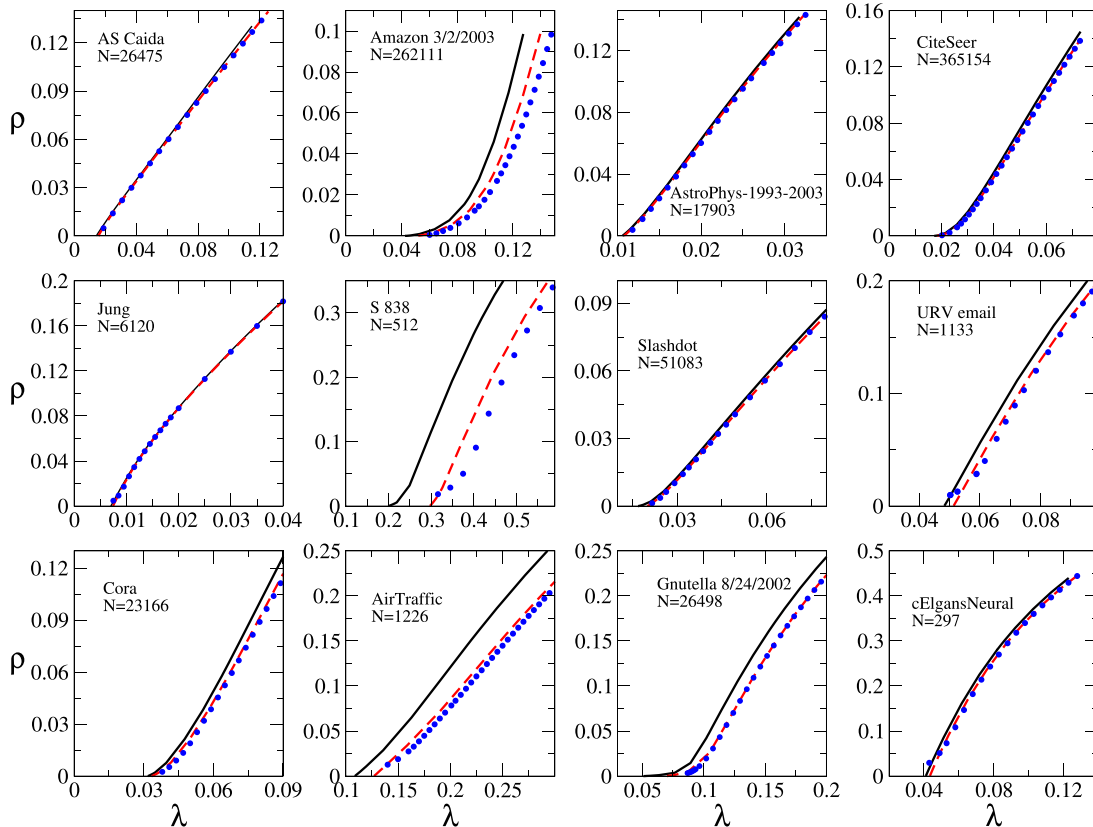


FIG. 4. Epidemic prevalence on real networks. Symbols represent stochastic simulations, while solid and dashed lines represent the numerical integration of the QMF and PQMF equations, respectively. In each panel, the usual name and size of the networks are given.

TABLE I. Relative deviations and inverse participation ratios for QMF and PQMF theories applied to real networks.

Network	η_{QMF}	η_{PQMF}	$Y_{\text{QMF}}^{(1)}$	$Y_{\text{PQMF}}^{(1)}$
AS Caida	0.032	0.0022	0.0240	0.0139
Amazon 3/2/03	0.75	0.24	0.106	0.0114
AstroPhys 93-03	0.035	0.013	0.0045	0.0043
CiteSeer	0.067	0.010	0.0177	0.0109
Jung	0.0079	0.0021	0.0478	0.0335
S 838	0.28	0.070	0.179	0.0340
Slashdot	0.032	0.0011	0.144	0.0347
URV email	0.022	0.0028	0.0096	0.0087
Cora	0.152	0.037	0.0100	0.0090
Air Traffic	0.38	0.066	0.0191	0.0154
Gnutella 8/24/02	0.17	0.0066	0.214	0.0800
cElegans Neural	0.057	0.0078	0.0189	0.0175

Amazon customers, electronic circuit S838, and Air Traffic networks, with 24%, 7%, and 6.6% of deviation, respectively.

C. Accuracy versus structural properties

The gain of PQMF theory with respect to QMF in real networks is expressive but it still deviates from simulations in some cases, as shown in Table I. One central question is to determine when either QMF or PQMF performance is satisfactory. Near the transition point, when the prevalence is very small, a relation between the accuracy and the localization of the PEV of the weighted adjacency matrices obtained in the linearization has been proposed [28]. This is justified by the fact that a leading contribution to the probability that a vertex i is infected in the mean-field theories is proportional to the corresponding PEV of A_{ij} or $B_{ij}(\lambda_c)$ for QMF and PQMF theories, respectively. For the sake of completeness of Ref. [28], in which the accuracy at the epidemic threshold was discussed thoroughly, Fig. 5 shows the steady-state density

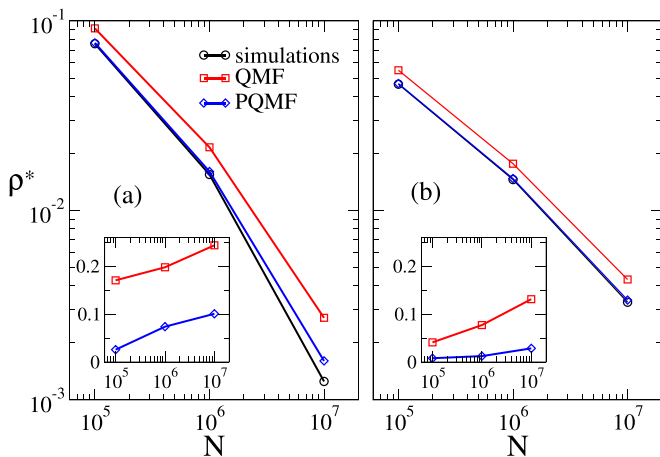


FIG. 5. Finite-size scaling of the steady-state density evaluated at $\lambda = 2\lambda_c^{\text{PQMF}}$. Weber-Porto configuration model networks with degree exponent $\gamma = 2.8$ presenting (a) disassortative ($\alpha = -0.2$) and (b) neutral ($\alpha = 0$) degree correlations are considered. Insets: Corresponding IPRs calculated for the PEV of the corresponding mean-field theory.

calculated slightly above the epidemic threshold of the PQMF theory versus the network size. The PQMF theory is much more accurate than the QMF but also starts to deviate from simulations as the network size increases. In both cases the accuracy is greater for networks with a less localized PEV as quantified by the inverse participation ratio (IPR) [29], defined as

$$Y^{(1)} = \sum_{i=1}^N [v_i^{(1)}]^4, \quad (10)$$

where $v_i^{(1)}$ is the normalized PEV defined in Sec. II. The larger the IPR, the more localized the PEV. The insets in Fig. 5 show the IPR for both QMF and PQMF theories where we see that the latter is much less localized than the former but still increases towards a finite value as the network size increases, indicating localization asymptotically.

However, the nonperturbative theory accounts for the contributions of the complete basis of eigenvectors, whether it is of A_{ij} or $B_{ij}(\lambda_c)$, and this comparison is no longer justifiable. Indeed, as shown in Table I, the accuracy of the QMF theory can be high even when the PEV is localized, as, for example, in the case of the Slashdot network. In other cases, such as the Air Traffic and Cora networks, the PEV localization corresponding to the QMF and PQMF theories are similar but the performance of the latter is much better. We performed a statistical analysis of the correlations between $\ln \eta$ and $\ln Y^{(1)}$ and found non-statistically significant p values of $p_{\text{QMF}} = 0.29$ and $p_{\text{PQMF}} = 0.46$. It is noteworthy that the statistical analyses considering the linear data present even lower statistical significance.

We also checked (logarithm) statistical correlations of η with other basic network metrics, namely, the heterogeneity coefficient $\varepsilon = \langle k^2 \rangle / \langle k \rangle$, the modularity coefficient Q [57], the average clustering coefficient $\langle c \rangle$, and the average shortest distances $\langle \ell \rangle$. We found statistical significance, at $p < 0.02$, only with ε and $\langle \ell \rangle$. Actually, ε and $\langle \ell \rangle$ are correlated since more heterogeneous networks tend to have shorter average distances due to the shortcuts introduced by hubs [52]. The correlation between η and $\langle \ell \rangle$ actually is not very surprising since one intuitively expects that the shorter the distances, the more accurate the mean-field hypothesis of neglecting long-range correlations becomes. One interesting feature is that the approximation given by Eq. (3) in the PQMF theory discards the possibility of triangles [27], in which the neighbors i and l of j are also connected. So, one could expect a worse performance in networks with a high clustering coefficient, but no statistical correlation with this metric was found ($p_{\text{PQMF}} = 0.51$). In summary, we could not infer which structural properties control the accuracy of the mean-field approaches in regimes of high prevalence.

IV. CONCLUSIONS

Theoretical understanding of dynamical processes on networks constitutes a powerful tool for protection against threats such as disease dissemination, misinformation propagation, and transportation infrastructure overload, among many other examples. Reliable theoretical approximations are usually required to consider the heterogeneous structure of the contact

networks and the dynamical correlations, in which the states of neighboring individuals are statistically correlated. These features are explicitly included in the PQMF theory [27]. However, this theory has been applied mainly to analyze the behavior of epidemic processes in the neighborhood of the transition from an endemic to a disease-free state through perturbative analyses where the epidemic prevalence is very low. In this work, we contribute to filling this gap by performing a detailed nonperturbative numerical analysis of the SIS model on synthetic and real networks within a wide range of heterogeneities and assortativities.

For synthetic networks generated with the Weber-Porto [37] configuration model, we report that the PQMF theory predicts with great accuracy the regime of high prevalence observed in stochastic simulations in networks with power-law degree distributions for all values of degree exponents investigated ($\gamma = 2.3, 2.8$, and 3.5) and degree correlations (disassortative, neutral, and assortative). In the case of a large $\gamma > 2.5$, where hubs tend to become separated as the network size increases, we observed that the PQMF theory significantly outperforms the simpler QMF theory where heterogeneity is fully considered but dynamical correlations are neglected, the discrepancy between theories being larger for large γ . The high accuracy of the PQMF theory at high prevalences contrasts with its bad performance at asymptotically low densities, where the theory is known to deviate from exactly known critical behavior [38], where $\rho \sim \lambda^\beta$ with $\beta > 1$ while the mean-field exponent is $\beta_{\text{MF}} = 1$ [28]. We argue, however, that this mismatch is constrained to a region very close to $\lambda = \lambda_c \rightarrow 0^+$ such that the regime of not too low density can still be accurately described by the PQMF theory.

In a set of real networks, where much more complex structures and correlations can be present, we observed that the PQMF theory always outperforms (sometimes very significantly) the QMF theory but may still present nonnegligible deviations from simulations in some cases (see Table I). Differently from the low-prevalence regime, where the accuracy of mean-field theories is correlated with the spectral properties of Jacobian matrices, only trivial statistical correlations with simple network metrics could be identified and the problem of predicting when nonperturbative analysis is sufficiently accurate given certain network properties remains open.

We expect that our work will stimulate the application of nonperturbative approaches through the numerical integration of continuous-time equations to address other fundamental problems of dynamical processes on networks. As perspective it would be interesting to perform a microscopic analysis of the validity of Eq. (3) at the triplet level. Moreover, the systematic comparison of nonperturbative mean-field theories and simulations could be extended to dynamical processes of a different nature such as the non-Markovian case [58].

ACKNOWLEDGMENTS

S.C.F. was partially supported by the Brazilian agencies CNPq (Grant No. 430768/2018-4) and FAPEMIG (Grant No. APQ-02393-18). F.A.R. acknowledges CNPq (Grant No. 309266/2019-0) and FAPESP (Grants No. 2016/25682-5 and No. 2013/07375-0) for the financial support given to this

research. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES), Finance Code 001.

APPENDIX A: WEBER-PORTO CONFIGURATION MODEL

The Weber-Porto configuration model networks are generated as follows. The degree of each vertex is drawn according to the degree distribution $P(k)$ such that each node has k unconnected half-edges. Two half-edges are chosen and connected with probability

$$P_{\text{link}}(q', q) = \frac{f(q', q)}{f_{\text{max}}}, \quad (\text{A1})$$

where q and q' are the respective degrees of the chosen vertices and f_{max} is the maximum value of

$$f(q, q') = 1 + \frac{(\kappa_{\text{nn}}(q) - \langle k \rangle_e)(\kappa_{\text{nn}}(q') - \langle k \rangle_e)}{\langle k \kappa_{\text{nn}} \rangle_e - \langle k \rangle_e^2}, \quad (\text{A2})$$

where $\langle A(k) \rangle_e = \sum_k A(k)P_e(k)$ and $P_e(k) = kP(k)/\langle k \rangle$ is the probability that an edge ends on a vertex of degree k . Self and multiple connections are forbidden. In the absence of degree correlations, we have $\kappa_{\text{nn}} = \langle k \rangle_e$, implying $f(q, q') = 1$ and $P_{\text{link}} = 1$. See Ref. [37] for more details.

APPENDIX B: QUASISTATIONARY METHOD

We applied the standard quasistationary method [44,45,59], where the dynamics returns to a previously visited active configuration with at least one infected vertex every time the system falls into the absorbing state where all vertices are simultaneously susceptible. The method is implemented by building and constantly updating a list with $M = 100$ active configurations. Every time the system falls into the absorbing state one of the M configurations is chosen with an equal chance and replacing the absorbing state. The list is updated with probability 10^{-2} by unit of time and the update consists of replacing a randomly selected configuration of the list by the present state of the dynamics. The quasistationary averages are computed during a time varying from $t_{\text{av}} = 10^5 \mu^{-1}$ to $t_{\text{av}} = 10^6 \mu^{-1}$ after a relaxation time $t_{\text{rlx}} = 10^5 \mu^{-1}$. Longer times are used for lower densities where fluctuations are more relevant.

APPENDIX C: STOCHASTIC SIMULATION OF THE SIS MODEL

Simulations of the SIS model were performed using the optimized Gillespie algorithm described in [44]. The number of infected vertices N_{inf} and the total number of edges emanating from them N_{SI} are computed and kept updated along the simulations. At each time step, with probability

$$q = \frac{\mu N_{\text{inf}}}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}, \quad (\text{C1})$$

one infected vertex is chosen with equal chance and healed. With the complementary probability

$$1 - q = \frac{\lambda N_{\text{SI}}}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}, \quad (\text{C2})$$

one infected vertex i is chosen with a probability proportional to its degree. One of the nearest neighbors of i , represented by j , is chosen with equal chance. If j is susceptible, it becomes infected, and otherwise, no change of state is implemented. The time is incremented by

$$\delta t = \frac{-\ln u}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}, \quad (\text{C3})$$

where u is a pseudorandom number uniformly distributed in the interval $(0,1)$.

-
- [1] I. Kiss, J. Miller, and P. Simon, *Mathematics of Epidemics on Networks: From Exact to Approximate Models* (Springer International, Cham, Switzerland, 2017).
- [2] W. Wang, M. Tang, H. Eugene Stanley, and L. A. Braunstein, Unification of theoretical approaches for epidemic spreading on complex networks, *Rep. Prog. Phys.* **80**, 036603 (2017).
- [3] 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).
- [4] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models. Aléa-Saclay* (Cambridge University Press, Cambridge, UK, 2005).
- [5] D. ben-Avraham and J. Köhler, Mean-field(n, m)-cluster approximation for lattice models, *Phys. Rev. A* **45**, 8358 (1992).
- [6] J. Joo and J. L. Lebowitz, Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice, *Phys. Rev. E* **70**, 036114 (2004).
- [7] J. P. Gleeson, High-Accuracy Approximation of Binary-State Dynamics on Networks, *Phys. Rev. Lett.* **107**, 068701 (2011).
- [8] J. P. Gleeson, Binary-State Dynamics on Complex Networks: Pair Approximation and Beyond, *Phys. Rev. X* **3**, 021004 (2013).
- [9] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, New York, 2008).
- [10] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.* **87**, 925 (2015).
- [11] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [12] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings* (IEEE Computer Society, New York, 2003), pp. 25–34.
- [13] K. T. D. Eames and M. J. Keeling, Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases, *Proc. Natl. Acad. Sci. USA* **99**, 13330 (2002).
- [14] M. J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proc. R. Soc. London Ser. B Biol. Sci.* **266**, 859 (1999).
- [15] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, Heterogeneous pair-approximation for the contact process on complex networks, *New J. Phys.* **16**, 053006 (2014).
- [16] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, The Kuramoto model in complex networks, *Phys. Rep.* **610**, 1 (2016).
- [17] S. Chatterjee and R. Durrett, Contact processes on random graphs with power law degree distributions have critical value 0, *Ann. Prob.* **37**, 2332 (2009).
- [18] T. Mountford, J.-C. Mourrat, D. Valesin, and Q. Yao, Exponential extinction time of the contact process on finite graphs, *Stoch. Process. Appl.* **126**, 1974 (2016).
- [19] C. Castellano and R. Pastor-Satorras, Thresholds for Epidemic Spreading in Networks, *Phys. Rev. Lett.* **105**, 218701 (2010).
- [20] C. Castellano and R. Pastor-Satorras, Competing activation mechanisms in epidemics on networks, *Sci. Rep.* **2**, 24 (2012).
- [21] 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).
- [22] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, Collective versus hub activation of epidemic phases on networks, *Phys. Rev. E* **93**, 032314 (2016).
- [23] W. Cota, A. S. Mata, and S. C. Ferreira, Robustness and fragility of the susceptible-infected-susceptible epidemic models on complex networks, *Phys. Rev. E* **98**, 012310 (2018).
- [24] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, Epidemic thresholds in real networks, *ACM Trans. Info. Syst. Secur.* **10**, 1 (2008).
- [25] M. Boguñá and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, *Phys. Rev. E* **66**, 047104 (2002).
- [26] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results, *Phys. Rev. E* **86**, 041125 (2012).
- [27] 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).
- [28] D. H. Silva, S. C. Ferreira, W. Cota, R. Pastor-Satorras, and C. Castellano, Spectral properties and the accuracy of mean-field approaches for epidemics on correlated power-law networks, *Phys. Rev. Res.* **1**, 033024 (2019).
- [29] 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).
- [30] 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).
- [31] J. T. Matamalas, A. Arenas, and S. Gómez, Effective approach to epidemic containment using link equations in complex networks, *Sci. Adv.* **4**, eaau4212 (2018).
- [32] Q. Wu and T. Hadzibeganovic, Pair quenched mean-field approach to epidemic spreading in multiplex networks, *Appl. Math. Model.* **60**, 244 (2018).
- [33] Q. Wu, R. Zhou, and T. Hadzibeganovic, Conditional quenched mean-field approach for recurrent-state epidemic dynamics in complex networks, *Phys. A Stat. Mech. Appl.* **518**, 71 (2019).

- [34] Q. Wu and T. Hadzibeganovic, An individual-based modeling framework for infectious disease spreading in clustered complex networks, *Appl. Math. Model.* **83**, 1 (2020).
- [35] 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).
- [36] P. G. Fennell, S. Melnik, and J. P. Gleeson, Limitations of discrete-time approaches to continuous-time contagion dynamics, *Phys. Rev. E* **94**, 052125 (2016).
- [37] S. Weber and M. Porto, Generation of arbitrarily two-point-correlated random networks, *Phys. Rev. E* **76**, 046111 (2007).
- [38] T. Mountford, D. Valesin, and Q. Yao, Metastable densities for the contact process on power law random graphs, *Electron. J. Probab.* **18**, 1 (2013).
- [39] P. Van Mieghem, The N-intertwined SIS epidemic network model, *Computing* **93**, 147 (2011).
- [40] P. Van Mieghem, The viral conductance of a network, *Comput. Commun.* **35**, 1494 (2012).
- [41] Obviously, this region will not shrink in a logarithm scale and the asymptotic scaling is the theoretical one given by Eq. (7).
- [42] The solution is the same one derived for the contact process in, e.g., Ref. [43], replacing the infection rate $\lambda_{CP} = m\lambda_{SIS}$, where λ_{CP} is the infection rate for the contact process and λ_{SIS} the infection rate for SIS.
- [43] R. S. Ferreira and S. C. Ferreira, Critical behavior of the contact process on small-world networks, *Eur. Phys. J. B* **86**, 462 (2013).
- [44] W. Cota and S. C. Ferreira, Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks, *Comput. Phys. Commun.* **219**, 303 (2017).
- [45] R. S. Sander, G. S. Costa, and S. C. Ferreira, Sampling methods for the quasistationary regime of epidemic processes on regular and complex networks, *Phys. Rev. E* **94**, 042308 (2016).
- [46] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Dynamical and Correlation Properties of the Internet, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [47] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Generation of uncorrelated random scale-free networks, *Phys. Rev. E* **71**, 027103 (2005).
- [48] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [49] A. S. Mata and S. C. Ferreira, Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks, *Phys. Rev. E* **91**, 012816 (2015).
- [50] R. Juhász, G. Ódor, C. Castellano, and M. A. Muñoz, Rare-region effects in the contact process on networks, *Phys. Rev. E* **85**, 066125 (2012).
- [51] M. M. de Oliveira, S. G. Alves, and S. C. Ferreira, Dynamical correlations and pairwise theory for the symbiotic contact process on networks, *Phys. Rev. E* **100**, 052302 (2019).
- [52] A.-L. Barabási and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, UK, 2016).
- [53] C. Castellano and R. Pastor-Satorras, Relating Topological Determinants of Complex Networks to Their Spectral Properties: Structural and Dynamical Effects, *Phys. Rev. X* **7**, 041024 (2017).
- [54] R. Pastor-Satorras and C. Castellano, Eigenvector localization in real networks and its implications for epidemic spreading, *J. Stat. Phys.* **173**, 1110 (2018).
- [55] F. Radicchi, Predicting percolation thresholds in networks, *Phys. Rev. E* **91**, 010801(R) (2015).
- [56] F. Radicchi and C. Castellano, Breaking of the site-bond percolation universality in networks, *Nat. Commun.* **6**, 10196 (2015).
- [57] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
- [58] M. Feng, S.-M. Cai, M. Tang, and Y.-C. Lai, Equivalence and its invalidation between non-Markovian and Markovian spreading dynamics on complex networks, *Nat. Commun.* **10**, 3748 (2019).
- [59] M. M. de Oliveira and R. Dickman, How to simulate the quasistationary state, *Phys. Rev. E* **71**, 016129 (2005).