

Cascades on correlated and modular random networks

James P. Gleeson

Department of Mathematics and Statistics, University of Limerick, Ireland

(Received 26 November 2007; published 28 April 2008)

An analytical approach to determining the mean avalanche size in a broad class of dynamical models on random networks is introduced. Previous results on percolation transitions and epidemic sizes are shown to be special cases of the method. The time-dependence of cascades and extensions to networks with community structure or degree-degree correlations are discussed. Analytical results for the rate of spread of innovations in a modular network and for the size of k cores in networks with degree-degree correlations are confirmed with numerical simulations.

DOI: [10.1103/PhysRevE.77.046117](https://doi.org/10.1103/PhysRevE.77.046117)

PACS number(s): 89.75.Hc, 64.60.aq, 87.23.Ge, 05.10.-a

I. INTRODUCTION

Network models are used to examine the underlying structure of complex systems ranging from the Internet to patterns of social interaction; see the reviews [1–3] and references therein. In certain circumstances, interactions between the nodes (vertices) of the network may cause initially localized effects to propagate throughout the whole network. Such *avalanches* or *cascades* occur, for example, in the transmission of infectious diseases through communities and of computer viruses over email networks. The adoption of innovations and the spread of fads may be modeled as cascade processes on social networks, with the aim of viral marketing being the creation of global cascades on these networks.

The dynamics of cascades are strongly dependent upon the topological structure of the underlying network and on the details of how the cascade spreads among the nodes of the network [4–15]. In the class of examples considered in this paper, each node of the network can be in one of two states: either *active* (also termed *damaged* or *infected*) or *inactive* (*undamaged* or *susceptible*), with nodes updating their states depending on the number and state of the node's immediate neighbors in the (undirected) network. Networks are chosen from an ensemble of graphs with specified degree distribution (i.e., using the configuration model [16]), and both synchronous and asynchronous updating may be considered. In this paper we show that for a class of such models the average cascade size may be determined analytically (averages being taken over an ensemble of realizations). This basic model is also extended to networks with strong community structure or with degree-degree correlations. Previous results on percolation and k -core sizes are shown to be special cases of our general approach.

We consider undirected networks of N binary-valued nodes in the $N \rightarrow \infty$ limit, where the probability of a node being activated depends only on its degree k and the number m of its neighbors who are already active. Denoting this probability by $F(m, k)$ (termed the *neighborhood influence response function* [17,18]) we require that (i) for any fixed k , $F(m, k)$ is nondecreasing with m , and (ii) once active, a node cannot become deactivated. These requirements mean that increasing the number of active neighbors of a given node will increase (or at least not decrease) the probability of the

chosen node itself becoming activated. This type of positive feedback mechanism (known in the sociological literature as “positive externalities”) ensures that the number of active nodes in a given realization is nondecreasing with time, and enables us to use analytical methods to calculate the mean avalanche size on locally treelike (vanishing clustering) random networks.

A broad range of dynamical problems on random undirected networks can be shown (see Sec. II) to obey conditions (i) and (ii), including Watts' model of threshold dynamics [13], k -core size calculations [19], site and bond percolation problems [20], and (under certain types of external field) zero-temperature random-field Ising models [21]. Susceptible-infected-recovered (SIR) disease transmission models may be mapped to the bond percolation problem (in steady state) [22], and so are also included in the class of problems solvable using our methods. Methods for calculating avalanche sizes on directed networks have also been examined recently [23]; here we restrict our attention to undirected networks.

Our analytical approach builds on methods introduced to study the zero-temperature random-field Ising model [21,24,25]. The main assumption is that the network may be approximated by a tree structure, and so networks with non-zero clustering (i.e., with short loops or cycles in the graph) or disconnected segments may not be considered. Generating function methods have been used to determine expected sizes of clusters (cascades) in percolation problems, and can in fact give the whole distribution of cascade sizes [1,16,20]. However, the generating function approach is not directly applicable to the wider class of models we consider here. Moreover, we believe the relative simplicity of our method makes the calculation of mean cascade sizes more intuitive even in those problems where generating functions have been applied.

The remainder of this paper is structured as follows. In Sec. II we show that an analytical approach introduced to find the mean cascade size in Watts' model of threshold dynamics may easily be generalized to a wide class of network problems exhibiting cascades. Connections to existing results on percolation, giant components, and k cores are highlighted. The time dependence of mean cascade sizes for both synchronous and asynchronous updating is considered in Sec. III. Extensions of the basic theory to model networks with community structure or degree-degree correlations are

derived in Secs. IV and V, and examples of time-dependent infection in community-structure networks and k -core sizes in correlated graphs demonstrate the excellent agreement of theory with numerical simulations. Appendixes A and B address the effects of single-seed activations, and the simplification of the equations under special assumptions.

II. GENERALIZED MODEL

In this section we briefly review our earlier work on Watts' model, and then demonstrate how analytical results for the mean cascade size may also be found for a variety of other cases of interest. The response functions introduced here form the basis for extending the theory to time-dependent cases on modular or correlated networks in Secs. III–V.

A. Watts' model

In our recent paper [26] we derived analytical expressions for the mean avalanche size in Watts' model of threshold dynamics [13]. In this model, each node of the network is assigned a random (frozen) threshold r from a specified distribution, and when updated, the node (of degree k , say) becomes active if the fraction m/k of its neighbors which are active exceeds r . Cascades are initiated by randomly activating a *seed fraction* ρ_0 of the nodes.

By approximating the random network by a tree structure and then defining the probability q_n that a random node at level n of the tree is active, conditional on its parent in the tree being inactive, we derived the following iteration equation:

$$q_{n+1} = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1 - q_n)^{k-1-m} F(m, k) \equiv g(q_n). \quad (1)$$

Here p_k is the degree distribution (probability that a node has k neighbors) of the configuration-model network, z is the mean degree $\sum_k k p_k$, and the response function is

$$F(m, k) = C\left(\frac{m}{k}\right), \quad (2)$$

where C is the cumulative distribution function (cdf) of the thresholds. If, for example, all nodes have the same threshold R , the response function is

$$F(m, k) = \begin{cases} 0 & \text{if } m \leq Rk, \\ 1 & \text{if } m > Rk. \end{cases} \quad (3)$$

The derivation of Eq. (1) uses the fact that the degree distribution of nearest neighbors on a tree is $k p_k / z$, and also that a node with $k-1$ n -level children of which m are active becomes active itself with probability $\binom{k-1}{m} q_n^m (1 - q_n)^{k-1-m} F(m, k)$. As discussed in detail in [26], it is crucial that the state of each node may be altered at most once—it is this property that allows us to ignore propagation of activity away from the central node of the tree and treat each node as activated from its children.

Given an initial random fraction ρ_0 of active nodes, we iterate Eq. (1) from $q_0 = \rho_0$ to convergence to determine $q_\infty = \lim_{n \rightarrow \infty} q_n$, and then find the expected steady-state density of active nodes in the network as

$$\rho = \rho_0 + (1 - \rho_0) \sum_{k=0}^{\infty} p_k \sum_{m=0}^k \binom{k}{m} q_\infty^m (1 - q_\infty)^{k-m} F(m, k) \equiv h(q_\infty). \quad (4)$$

To derive Eq. (4) we examine the central node of the tree, which has k children with probability p_k [26]. The quantity ρ is the average (over an ensemble of realizations) of the steady-state cascade size in networks characterized by p_k , with dynamics specified by $F(m, k)$.

The only feature of the response function F that is crucial to our derivations is that it is a nondecreasing function of m for any fixed k . It is therefore straightforward to generalize our approach to calculate the expected cascade size in a variety of dynamical problems which also obey this condition (and in which nodes, once activated, remain permanently active). Note that the use of the tree structure in this derivation assumes that the fraction ρ_0 of seed nodes is nonvanishing. Cases where cascades are initialized by single-node seeds (i.e., $\rho_0 = 1/N$) require further consideration; see Appendix A.

B. Other models

1. Absolute number of active neighbors

In their recent paper [27], Galstyan and Cohen examine a version of Watts' model in which the absolute number of active neighbors, rather than the fraction of the total number as in [13], is compared with the nodes' thresholds to determine the subsequent states. The corresponding response function for Eqs. (1) and (4) is thus

$$F(m, k) = C(m), \quad (5)$$

with C being the cdf of thresholds as before. We consider this example in some detail in Sec. IV below, where the fact that the response function is independent of the nodal degree k will lead to significant simplification.

2. Site and bond percolation

Percolation on random networks has been extensively studied [20, 28–30] and formulas for the size of the giant connected component have been determined using generating function methods [20] and applied to study network resilience and epidemic thresholds. Here we point out that these are special cases of our general approach, corresponding to suitable choices for the response function in Eqs. (1) and (4).

In the bond percolation problem, network edges are *occupied* with probability p and nodes become *infected* (active) if they are linked to an infected node by an occupied edge. Thus with m active neighbors, a node has probability $(1-p)^m$ of not becoming infected, so the response function for this case is

$$F(m,k) = \begin{cases} 0 & \text{if } m = 0, \\ 1 - (1-p)^m & \text{if } m > 0. \end{cases} \quad (6)$$

Taking the $\rho_0 \rightarrow 0$ limit in Eqs. (1) and (4) reproduces the equations for the size of the giant component (epidemic size) [20] given, for example, in Eqs. (8.11) of [1], by applying the following result for the response function (6):

$$\sum_{m=0}^k \binom{k}{m} q^m (1-q)^{k-m} F(m,k) = 1 - (1-pq)^k. \quad (7)$$

For site percolation, nodes of degree k are occupied with probability Q_k and occupied nodes become active if they have one or more active neighbors. Unoccupied nodes can never become active. The response function for site percolation is therefore

$$F(m,k) = \begin{cases} 0 & \text{if } m = 0, \\ Q_k & \text{if } m > 0. \end{cases} \quad (8)$$

Using this in the $\rho_0 \rightarrow 0$ limit of Eqs. (1) and (4), and noting that for this response function we have

$$\sum_{m=0}^k \binom{k}{m} q^m (1-q)^{k-m} F(m,k) = Q_k [1 - (1-q)^k], \quad (9)$$

we reproduce Eqs. (8.5) of [1] for the size of the giant component.

3. k -core sizes

The k core of a network is the largest subgraph whose nodes have degree at least k [31]. Study of k -core decompositions gives insights into the topology of interconnected parts of real-world networks such as the Internet [32]. Recently analytical descriptions of the k -core sizes were found for random uncorrelated networks [19,31]. As discussed in [19], the size of the k core may be calculated as the steady state of a cascade process—at each iteration all nodes with less than k (undamaged) neighbors are removed from the network (or, equivalently, are labeled as damaged themselves). In the steady-state limit of this cascade process, only nodes in the k core remain undamaged.

To calculate the k -core size for $k=K$, we label damaged nodes as active (i.e., “active” nodes=damaged or removed nodes), and allow an initial random fraction ρ_0 to be damaged. A node of degree k becomes damaged if the number m of its damaged neighbors is greater than $k-K$, i.e., if less than K of its neighbors are undamaged. Hence we set the response function to be

$$F(m,k) = \begin{cases} 0 & \text{if } m \leq k - K, \\ 1 & \text{if } m > k - K \end{cases} \quad (10)$$

in Eq. (1).

The final density ρ of damaged nodes then gives a size $1-\rho$ for the undamaged k core for $k=K$ as a fraction of the total (undamaged) network size. It is straightforward to confirm that the k -core size in uncorrelated networks (as given in Eqs. (1) and (2) of [19]) is given by our Eqs. (1) and (4) using Eq. (10) as above. In Sec. V we extend the method to

find interesting results on k -core sizes in networks with degree-degree correlations.

C. Cascade condition

For all the examples considered above, a *cascade condition* may be derived which determines whether an infinitesimally small seed fraction ρ_0 of activated nodes will generate a nonvanishing mean cascade size. The criterion is that the iteration of Eq. (1) must cause an infinitesimally small value of $q_0=\rho_0$ to grow; examining the derivative of the right hand side of Eq. (1) with respect to q_n at $q_n=0$ gives the condition for cascades to occur:

$$\sum_k \frac{k}{z} (k-1) p_k [F(1,k) - F(0,k)] > 1. \quad (11)$$

Use of the response functions (6) or (8) in Eq. (11) yields well-known conditions for critical percolation previously derived using other approaches [33], including the Molloy-Reed criterion for existence of a giant connected component [34]. The cascade condition derived by Watts for his model of threshold dynamics [13] is also given by Eq. (11); note that for this case we have recently extended the cascade condition to cases where ρ_0 is small but nonvanishing [26]. In Secs. IV and V we will extend the cascade condition (11) to modular and correlated networks.

III. TIME-DEPENDENT SOLUTIONS

In this section we consider how the analytical solution (1) and (4) may be extended to give the time-dependence of the fraction ρ of active nodes as well as its steady-state value. Using the appropriate response function [e.g., Eqs. (2), (5), (6), (8), and (10)] immediately allows these results to apply also to the percolation, k -core, and other problems considered in Sec. II.

In the case where all N nodes are updated synchronously the tree level n used in the derivation of Eqs. (1) and (4) may be mapped directly to the time step n . Then the expected fraction q_{n+1} of nodes with inactive parents at time step $n+1$ depends on the value at the previous time step as given by the function g of Eq. (1):

$$q_{n+1} = g(q_n). \quad (12)$$

The total fraction of active nodes at time step $n+1$ is denoted by ρ_{n+1} and its value is given in terms of q_n as

$$\rho_{n+1} = h(q_n), \quad (13)$$

with h as defined in Eq. (4). Note that the steady-state solution of [26] given in Eq. (4) follows by taking the limit $n \rightarrow \infty$ in Eq. (13). The time dependence of the synchronously updated bond percolation problem has recently been considered in the context of epidemic dynamics [35].

So far we have considered only synchronous updating, i.e., all N network nodes are updated simultaneously at each time step. In various applications (such as [27]) it is appropriate instead to update only a (randomly chosen) fraction f of the N nodes at each time step. Thus the case $f=1/N$ cor-

responds to updating a single node at each time step (i.e., standard asynchronous updating), while $f=1$ reduces to the synchronous updating case studied above. It may be shown [36] that both synchronous and asynchronous updating lead to the same steady state results, but of course the transient behavior may be very different for the two cases.

For sufficiently low values of f the asynchronously updated nodes are so sparse in the network that they may be considered to be independent of each other, and the time dependence of the averaged network quantities may be approximated as continuous: we introduce the obvious notation $q(t)$ and $\rho(t)$. To understand this asynchronous limit we note that the fraction of nodes which would be activated if updating were synchronous is [from Eq. (1)] $g(q(t))$. Thus a fraction $g(q(t)) - q(t)$ (assuming this is non-negative) of the network nodes is available for activation, and a randomly selected fraction f of these is updated, leading to (for $f \ll 1$) the following evolution equation for $q(t)$:

$$\frac{dq(t)}{dt} = f[g(q(t)) - q(t)]^+, \quad (14)$$

with initial condition $q(0) = \rho_0$ and the notation $[\cdot]^+$ standing for $\max(0, \cdot)$.

The density of active nodes at time t is similarly denoted $\rho(t)$ and is well approximated by the solution of

$$\frac{d\rho(t)}{dt} = f[h(q(t)) - \rho(t)]^+ \quad (15)$$

with $\rho(0) = \rho_0$ and the function h as given in Eq. (4).

As shall be shown in the following sections, appropriate vector-valued generalizations of both the synchronous and asynchronous updating cases may be derived using similar arguments to those introduced here, and show good agreement with numerical simulations (see Fig. 2, for example).

IV. MODULAR NETWORKS

Motivated by sociological and epidemiological data, there has been significant recent research interest in the study of dynamics on networks composed of two or more *modules* (or *communities*) of nodes [27,37]. The nodes in each module share similar characteristics (e.g., having the same degree distribution, threshold distribution, or response function) but these may vary dramatically from community to community. Typically nodes forming a community have a higher density of links within the group than to outside communities. The communities are linked to each other to form the overall network, but global properties may depend strongly on the heterogeneity of the network. It is therefore of interest to contrast the dynamics on networks with this community structure with the more homogeneous networks of the previous sections.

The effects of modularity upon percolation on networks was examined in [37] and here we adopt the same model for the network topology while generalizing the dynamics to the variety of response functions outlined in Sec. II, for which the generating function approach of [37] is not generally applicable. We will also compare our results with those of the

specific bimodular network examined in [27].

Consider a network of N nodes divided into d communities labeled $i=1, 2, \dots, d$. Each node is a member of one community, and we will refer to the members of community i as “ i -type” nodes. The i -type nodes draw their degrees k from the community-specific degree distribution $p_k^{(i)}$; there may be a different distribution for each community. Each edge from an i -type node may link to a j -type node with a fixed probability. These probabilities are described conveniently by the $d \times d$ mixing matrix \mathbf{e} [37], with e_{ij} defined as the probability that a randomly chosen edge in the network connects a node of type i to a node of type j . We denote by $N^{(i)}$ the number of i -type nodes in the network, and note that \mathbf{e} will depend on the distribution of node types.

It is relatively straightforward to generalize the derivation of Eqs. (1) and (4) to a system of equations describing the modular case. Again, we stress that the conditions (i) and (ii) discussed in Sec. I (i.e., $F(m, k)$ nondecreasing with m and permanently activated nodes) are necessary requirements for our derivation to apply. Given an inactive i -type node and conditioning on its parent being inactive, the probability that each of its children is active at (synchronous) time step n is denoted by $\bar{q}_n^{(i)}$, with

$$\bar{q}_n^{(i)} = \frac{\sum_j e_{ij} q_n^{(j)}}{\sum_j e_{ij}}, \quad (16)$$

and the updating equation for each type takes the form

$$q_{n+1}^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_k \frac{k}{z^{(i)}} p_k^{(i)} \sum_{m=0}^{k-1} \binom{k-1}{m} (\bar{q}_n^{(i)})^m \times (1 - \bar{q}_n^{(i)})^{k-1-m} F^{(i)}(m, k), \quad (17)$$

with $q_0^{(i)} = \rho_0^{(i)}$. Here $z^{(i)}$ is the mean degree ($z^{(i)} = \sum k p_k^{(i)}$), $\rho_0^{(i)}$ the initial seed fraction, and $F^{(i)}(m, k)$ the response function, of the i -type nodes. As an obvious generalization of Eq. (12) we write this as

$$q_{n+1}^{(i)} = g^{(i)}(\bar{q}_n^{(i)}). \quad (18)$$

The density of active i -type nodes at time step $n+1$ is given by

$$\rho_{n+1}^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_k p_k^{(i)} \sum_{m=0}^k \binom{k}{m} (\bar{q}_n^{(i)})^m (1 - \bar{q}_n^{(i)})^{k-m} \times F^{(i)}(m, k) \equiv h^{(i)}(\bar{q}_n^{(i)}). \quad (19)$$

Finally, the total fraction of active nodes across the network is given by

$$\rho_n = \sum_i \frac{N^{(i)}}{N} \rho_n^{(i)}. \quad (20)$$

As in Sec. III these synchronous-update equations may be adapted for asynchronous updating, provided that the fraction f of nodes updated at each time step is sufficiently small. The resulting system of differential equations takes the form

$$\frac{dq^{(i)}(t)}{dt} = f[g^{(i)}(\bar{q}^{(i)}(t)) - q^{(i)}(t)]^+, \quad (21)$$

with

$$\bar{q}^{(i)}(t) = \frac{\sum_j e_{ij} q^{(j)}(t)}{\sum_j e_{ij}}, \quad (22)$$

and $q^{(i)}(0) = \rho_0^{(i)}$. The density equations are

$$\frac{d\rho^{(i)}(t)}{dt} = f[h^{(i)}(\bar{q}^{(i)}(t)) - \rho^{(i)}(t)]^+, \quad (23)$$

with $\rho^{(i)}(0) = \rho_0^{(i)}$, and the overall density is given by

$$\rho(t) = \sum_i \frac{N^{(i)}}{N} \rho^{(i)}(t). \quad (24)$$

Cascade conditions were discussed for the $d=1$ homogeneous network case in Sec. II C. The generalization to modular networks is most easily derived by considering a perturbation of the asynchronous-updating case about the steady state with $q^{(i)}=0$ for all i . Linearization of Eq. (21) using Eq. (22) yields the approximation

$$\frac{d\bar{q}^{(i)}(t)}{dt} \approx f \sum_j (A_{ij} - \delta_{ij}) \bar{q}^{(j)}(t), \quad (25)$$

with the entries of the $d \times d$ matrix \mathbf{A} given by

$$A_{ij} = \frac{e_{ij}}{\sum_\ell e_{i\ell}} \sum_k \frac{k}{z^{(j)}} (k-1) p_k^{(j)} [F^{(j)}(1,k) - F^{(j)}(0,k)]. \quad (26)$$

Infinitesimally small seed fractions may only grow to global cascades if at least one eigenvalue of the matrix $\mathbf{A} - \mathbf{I}$ is positive—equivalently, if an eigenvalue of \mathbf{A} is greater than 1. This matrix cascade condition is the natural generalization of Eq. (11) to the $d > 1$ case, and includes the phase transition condition for the emergence of the giant component found in [37] as a special case, using the bond percolation response function (6) in Eq. (26).

As an example of an asynchronously updated modular network we consider $d=4$ communities with links between them as denoted schematically in Fig. 1. Type 1 and type 2 nodes both have Poisson degree distributions with respective means $z^{(1)}=5.8$ and $z^{(2)}=8$, while type 3 and type 4 nodes have regular degree distributions, i.e., each node has exactly $z^{(3)}=z^{(4)}=8$ neighbors. The communities are all of equal size, so $N^{(i)}=N/4$ for $i=1-4$. Connections between the different types of nodes are quantified by the mixing matrix \mathbf{e} :

$$\mathbf{e} = \frac{1}{29.8} \begin{pmatrix} 5.5 & 0.15 & 0.15 & 0 \\ 0.15 & 7.7 & 0 & 0.15 \\ 0.15 & 0 & 7.7 & 0.15 \\ 0 & 0.15 & 0.15 & 7.7 \end{pmatrix}. \quad (27)$$

The zero elements of this matrix indicate that type 1 nodes have no direct link to type 4 nodes, and similarly there are no direct links between nodes in modules 2 and 3.

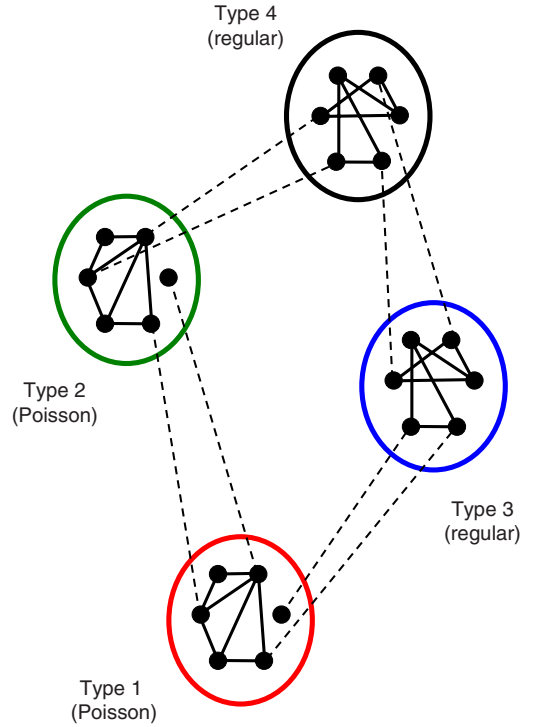


FIG. 1. (Color online) Schematic showing the $d=4$ communities with intermodule links described by the mixing matrix of Eq. (27).

We use asynchronous updating (with $f=0.01$) and initially activate only type 1 nodes, so $\rho_0^{(i)}=0$ for $i=2,3,4$, with $\rho_0^{(1)}=0.01$. It is interesting to observe the dynamics of the cascade process as the intercommunity links depicted in Fig. 1 facilitate the spread of the contagion. Here, unlike [27], we use Watts' original cascade model (3) with uniform threshold $R=0.18$.

Symbols show numerical results averaged over ten realizations with $N=5 \times 10^5$ nodes and curves are the outputs of Eqs. (21)–(24). In Fig. 2(a) we show the time-dependent density $\rho(t)$ of active nodes in the whole network. The rate of activation, $d\rho(t)/dt$, is plotted in Fig. 2(b), and clearly shows the effects of the community structure in the network. The ordering of the various community activations is highlighted in Fig. 2(c) by splitting the rate of activation into its four constituent parts $d\rho^{(i)}/dt$.

The initialization targets type 1 nodes, with this community (dashed curve) activated almost entirely before the other communities are appreciably affected. Although module 1 is linked equally strongly to both type 2 and type 3 nodes, the fact that these communities have different internal degree distributions causes module 2 (solid curve) to activate significantly earlier than module 3 (dash-dot curve). Type 3 nodes are then closely followed by type 4 (dotted) nodes, so that the final two peaks in Fig. 2(b) show relatively little separation. In all cases we find the theoretical results match the numerical simulations very well.

Perhaps surprisingly, if the initial activation targets type 4 nodes (instead of type 1 nodes as shown here) the cascade process does not spread beyond the tiny fraction of type 4 nodes and overall the network remains uninfected. This subtle effect is due to the different resiliences of the two

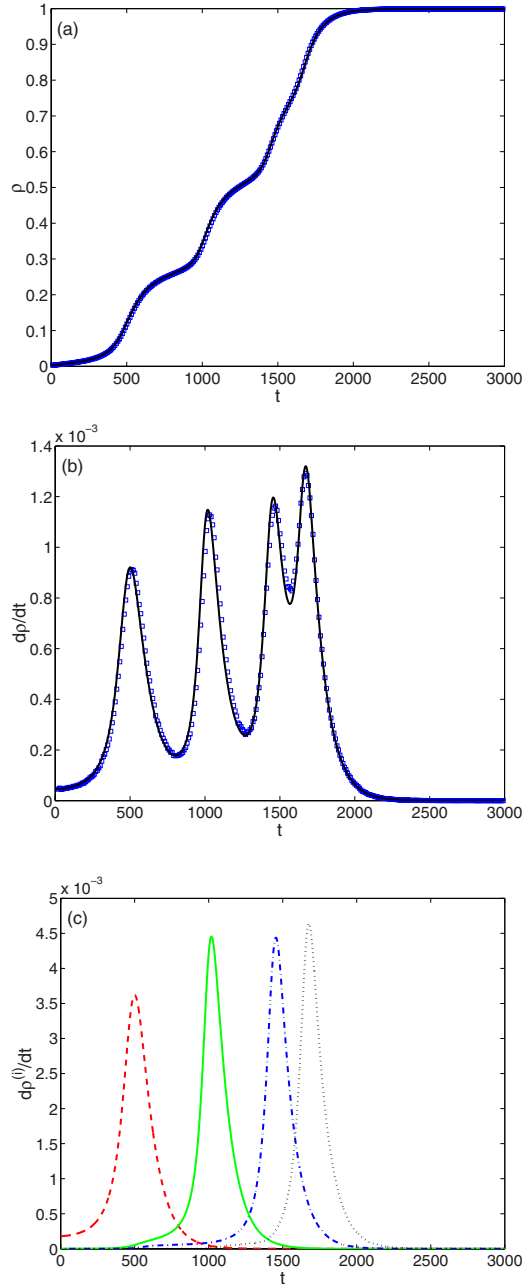


FIG. 2. (Color online) Numerical simulation results (symbols) and theoretical predictions of Eqs. (21)–(24) (curves) for Watts’ threshold dynamics model with uniform thresholds $R=0.18$ on the modular network shown schematically in Fig. 1. A fraction $f=0.01$ of all nodes are updated at each time step, and the cascade is initialized by a seed fraction $\rho_0^{(1)}=0.01$ of type 1 nodes. (a) Density of active nodes $\rho(t)$ for whole network; (b) rate of activation dp/dt in whole network; (c) rates of activation $dp^{(i)}/dt$ for each community: $i=1$ (dashed), $i=2$ (solid); $i=3$ (dash-dot), $i=4$ (dotted). Numerical simulations show average values over ten realizations of networks with $N=5 \times 10^5$ nodes.

communities, and may prove important in, for example, applications to viral marketing strategies. It is also interesting to note that the cascade condition on the matrix (26) is *not* satisfied in this example (the largest eigenvalue of \mathbf{A} is approximately 0.935), despite the fact that type 1 seed fractions

as low as $\rho_0^{(1)}=10^{-4}$ are observed to generate cascades. The linearization used in the derivation of Eq. (26) from the fully nonlinear equations is valid only for vanishingly small seed fractions. However, this example is one of the many realistic situations with small seed fractions where nonlinear effects are dominant, as demonstrated in [26] for the $d=1$ case.

Galstyan and Cohen [27] have recently considered cascade dynamics in networks composed of two modules, in the special case where each community has a Poisson degree distribution. In Appendix B we show that the general system of Eqs. (21)–(24) reduces to the simpler system of [27] provided that (i) each response function $F^{(i)}(m, k)$ is independent of k , and (ii) each community has a Poisson degree distribution.

V. DEGREE-DEGREE CORRELATIONS

To extend our analytical solutions beyond the case of uncorrelated networks we adopt the approach used in [37,38] for percolation problems on networks with degree-degree correlations and again extend to the wide class of possible dynamics described in Sec. II.

Let $P(k, k')$ be the joint probability distribution function (pdf) for the degrees k and k' of end vertices of a randomly chosen edge of the graph. In uncorrelated networks $P(k, k')$ factors as $k p_k k' p_{k'}/z^2$ since the degrees of nodes at either end of an edge are independent. In the more general case the Pearson correlation coefficient r may be defined as in [38]:

$$r = \frac{\sum_{k,k'} k k' P(k, k') - \left[\sum_{k,k'} k P(k, k') \right]^2}{\sum_{k,k'} k^2 P(k, k') - \left[\sum_{k,k'} k P(k, k') \right]^2}, \quad (28)$$

so that $r=0$ for the uncorrelated case. Networks with positive correlation $r \in (0, 1)$ are termed “assortative” and those with negative values $r \in (-1, 0)$ are termed “disassortative.” Newman has determined that r typically is positive in social interaction networks, while many technological networks such as the Internet (at autonomous systems level) show disassortative mixing [38].

Our analytical solution method may be extended to correlated networks in a similar fashion as for the modular networks seen in Sec. IV. Define $q_n^{(k)}$ as the probability, at (synchronous-updating) time step n , that a node of degree k is active, conditional on its parent being inactive. Similarly, writing $\bar{q}_n^{(k)}$ for the probability at time step n that a child of an inactive node of degree k is active, we have

$$\bar{q}_n^{(k)} = \frac{\sum_{k'} P(k, k') q_n^{(k')}}{\sum_{k'} P(k, k')} \quad (29)$$

[since a neighbor of the degree- k node has degree k' with probability $P(k, k')/\sum_{k'} P(k, k')$]. Then, similar to Eq. (17), the probabilities for each degree are updated as

$$q_{n+1}^{(k)} = \rho_0^{(k)} + (1 - \rho_0^{(k)}) \sum_{m=0}^{k-1} \binom{k-1}{m} (\bar{q}_n^{(k)})^m \times (1 - \bar{q}_n^{(k)})^{k-1-m} F(m, k), \quad (30)$$

where $\rho_0^{(k)}$ is the fraction of nodes of degree k which are

initially activated and initial conditions are $q_0^{(k)} = \rho_0^{(k)}$. The unconditional density of active k -degree nodes is given by

$$\rho_{n+1}^{(k)} = \rho_0^{(k)} + (1 - \rho_0^{(k)}) \sum_{m=0}^k \binom{k}{m} (\bar{q}_n^{(k)})^m (1 - \bar{q}_n^{(k)})^{k-m} F(m, k), \quad (31)$$

with the overall network density being equal to

$$\rho_n = \sum_k p_k \rho_n^{(k)}. \quad (32)$$

Corresponding equations for the case of asynchronous updating may be derived exactly as was done in Sec. IV.

The condition which allows cascades from infinitesimal seeds is that at least one eigenvalue of \mathbf{B} should exceed 1, where \mathbf{B} is defined as

$$B_{kj} = \frac{(j-1)}{\sum_{k'} P(k, k')} P(k, j) [F(1, j) - F(0, j)]. \quad (33)$$

This generalizes the $p=1$ bond percolation case (and giant component existence criterion) given in [38] to the broad class of applications with response functions listed in Sec. II.

An important example of the application of our method is the calculation of k -core sizes in networks with degree-degree correlations. As noted in Sec. II above, k cores in uncorrelated networks have been determined in [19,31], using an approach which may be viewed as an example of our generalized method for the specific response function (10). The approach of the current section is therefore applied to examine the effect of nonzero correlations upon k -core sizes, using the steady-state predictions of Eqs. (30)–(32), and noting that the final size of the k core is given by $1 - \rho_\infty$. Initial random damage of a fraction Q of nodes may be included in our calculations.

We concentrate on the $k=7$ core in a network with Poisson degree distribution of mean $z=10$. Degree-degree correlated networks are created as described in [33] by starting with pairs (k, k') of end-vertex degree values, drawn from an appropriate joint distribution, i.e., k and k' are the degrees of nodes at either end of a random edge in the graph. End vertices with the same k values are then collected together into random bunches of k nodes and identified with a single network node. When the network is sufficiently large (we use 5×10^5 edges, giving 10^5 nodes) we find that the effect of errors due to mismatched ends of links is negligibly small [37].

To generate the correlated pairs (k, k') we use a transformation method sometimes known as the Gaussian copula model [39]. For each pair we draw three numbers M , y_1 , and y_2 from independent zero-mean, unit-variance, Gaussian distributions and combine them to form correlated unit-variance Gaussian variables x_1 and x_2 as follows:

$$\begin{aligned} x_1 &= \sqrt{|\nu|} M + \sqrt{1 - |\nu|} |y_1|, \\ x_2 &= \text{sgn}(\nu) \sqrt{|\nu|} M + \sqrt{1 - |\nu|} |y_2|, \end{aligned} \quad (34)$$

where $\nu \in (-1, 1)$ is a copula correlation parameter, which is closely related to, though not identical to, the correlation

coefficient r of Eq. (28). Defining $G(x)$ as the cumulative distribution function (cdf) of x_1 (or x_2),

$$G(x) = \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{x}{\sqrt{2}}\right), \quad (35)$$

and $C(k)$ as the cdf for the degree of end-vertices of a random edge,

$$C(k) = \sum_{j=0}^k \frac{j}{z} p_j, \quad (36)$$

the values of (x_1, x_2) are transformed to the degrees (k, k') via

$$k = C^{-1}[G(x_1)], \quad k' = C^{-1}[G(x_2)]. \quad (37)$$

The joint pdf $P(k, k')$ of the degrees of end vertices is given by the double integral over the correlated joint-Gaussian distribution:

$$\begin{aligned} P(k+1, k'+1) &= \frac{1}{2\pi\sqrt{1-\nu^2}} \int_{G^{-1}[C(k)]}^{G^{-1}[C(k+1)]} \int_{G^{-1}[C(k')] }^{G^{-1}[C(k'+1)]} \\ &\quad \times \exp\left(-\frac{(x_1^2 + x_2^2 - 2\nu x_1 x_2)}{2(1-\nu^2)}\right) dx_1 dx_2, \end{aligned} \quad (38)$$

and may be used in Eqs. (30)–(32) to determine the theoretical results.

The correlation coefficient r resulting from a given value ν of the copula correlation parameter is calculated directly using Eqs. (28) and (38). For the cases considered here we find that the value of r is very close to the input value of ν ; specifically, for the values of $\nu = -0.5$ and 0.98 reported below we find $r = -0.49$ and 0.97 , respectively. Accordingly we quote only values of the copula correlation parameter ν in the following.

Figure 3 shows the effect of correlation values in the range $\nu \in (-1, 1)$ on the size of the undamaged ($Q=0$) $k=7$ core in the Poisson random graph with $z=10$. Compared to the uncorrelated case, it is clear that the size of the k core grows in disassortative networks, and decreases (with a sharp transition) for positive correlation values. The behavior near $\nu = \pm 1$ is rather pathological: as ν approaches 1, for example, the network fragments into subnetworks of regular degree in order to become completely assortative, so the k -core size approaches a limit of $\sum_{j=k}^{\infty} p_j$. The approach to this state as ν increases is through a series of plateaus which are well reproduced in both numerical simulations (symbols) and theoretical results (curve).

The dependence of the k -core size upon the initially damaged fraction Q of nodes is shown in Fig. 4. Assuming non-targeted damage [20] we take $\rho_0^{(k)} = Q$ for all degrees k and compare disassortative ($\nu = -0.5$, dotted line) and strongly assortative ($\nu = 0.98$, solid line) networks with the neutral ($\nu = 0$, dashed line) case. Again, the effects of degree-degree correlations upon the k -core size and upon the transitions observed is quite dramatic, while the agreement between theory and numerics is excellent. Equations (30)–(32) may be used to understand the complex relationships among the

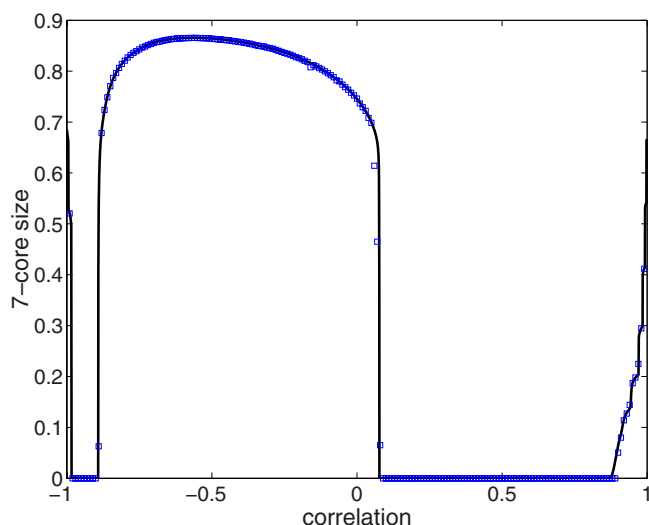


FIG. 3. (Color online) Average size of $k=7$ core in a $z=10$ undamaged Poisson random graph with degree-degree correlations, as a function of the copula correlation parameter ν . Symbols denote the average over ten realizations of networks with $N \approx 10^5$ nodes; the curve shows the theoretical prediction, using the matrix (38) in Eqs. (30)–(32).

parameters in correlated networks, and hopefully facilitate comparison with measurement of k cores in real-world networks [19], and assist in k -core-based modeling of the Internet [32].

VI. CONCLUSIONS

We have shown that a class of problems on networks, which includes Watts’ dynamical model [13] as well as several well-studied (static) network properties, such as perco-

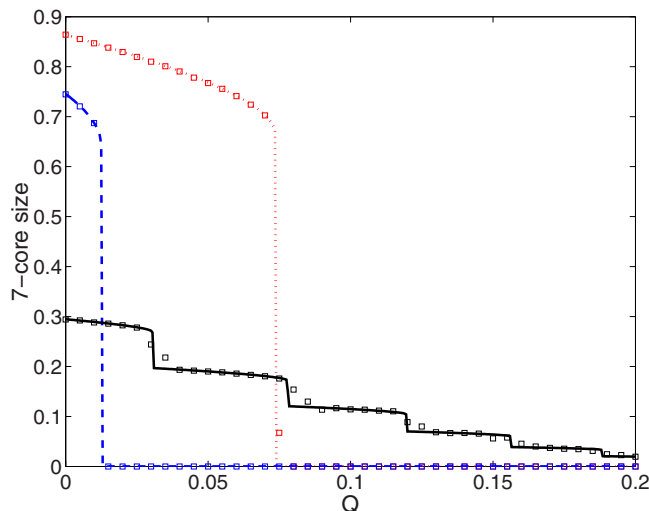


FIG. 4. (Color online) Size of $k=7$ cores in Poisson random graphs with $z=10$ as a function of initial damage Q , for copula correlation parameter $\nu=-0.5$ (dotted), $\nu=0$ (dashed), and $\nu=0.98$ (solid). Curves show theoretical predictions, and symbols denote numerical results over ten realizations of networks with $N \approx 10^5$ nodes.

lation transitions and k -core sizes, may be considered as special cases of a generalized model exhibiting cascade dynamics. The steady-state mean cascade size for this class of models is found analytically through Eqs. (1) and (4), and its time dependence under synchronous and asynchronous updating is considered in Sec. III. Further extensions to modular and degree-correlated networks (Secs. IV and V) are based upon network constructions introduced in Newman’s study [37] of percolation dynamics, which may be viewed as a special case of our model.

As well as providing a general framework for understanding several apparently different network problems, we have shown examples where our approach yields fresh insights. The spread of infection within Watts’ threshold dynamics model through a population of multiple distinct types has attracted recent interest [27], with existing results limited to two modules with Poisson degree distributions. Our analysis demonstrates how the general case of d types with arbitrary degree distributions may be solved; see Fig. 2. Our second example is the calculation of k -core sizes in networks with arbitrary degree-degree correlations (see Figs. 3 and 4), extending results previously limited to uncorrelated networks [19,31], and of special relevance for models of the Internet based on k cores [32].

We anticipate further study of cascade dynamics based on the analytical approach introduced here. In particular we highlight the interesting (numerical) results on the role of particularly influential seeds in a marketing context [18], the important role of k cores in the correlated network that is the Internet [32], and the possibility of extending our methods to include networks with nonzero clustering coefficients [40].

ACKNOWLEDGMENTS

This work was funded by Science Foundation Ireland under programmes 06/IN.1/I366 and 05/RFP/MAT0016. The author benefited from helpful discussions with Diarmuid Cahalane and Stephen Wills.

APPENDIX A: SINGLE-SEED INITIALIZATION

An implicit assumption of the derivations used in the main body of this paper (and in [26]) is that the initial seed fraction ρ_0 activates sufficiently many nodes for the tree structure used to be a valid approximation to the original network. For single seeds with $\rho_0=1/N$ (or other cases with $\rho_0 N$ relatively small) the probability of this approximation being invalid in any given realization is non-negligible. In Watts’ original paper [13] this single-seed effect was recognized and quantified in terms of the size S_v of the *vulnerable cluster*, and the size S_e of the *extended vulnerable cluster*.

The vulnerable cluster consists of all nodes which will activate if a single neighbor is activated, while the extended vulnerable cluster includes the vulnerable nodes, plus those nodes immediately adjacent to the vulnerable cluster. Thus a single-node seed will initiate a cascade (of expected size ρ as per the theory of Sec. II) if it is located in the extended vulnerable cluster but otherwise no global cascade will occur.

Taking the size of the extended vulnerable cluster as S_e and with $\lfloor \rho_0 N \rfloor$ initial seed nodes, the probability that none of these fall within the extended vulnerable cluster is $(1 - S_e)^{\lfloor \rho_0 N \rfloor}$. Therefore the probability that a cascade occurs in a given realization is $1 - (1 - S_e)^{\lfloor \rho_0 N \rfloor}$ and the average size of occurring cascades is ρ , so a (rather crude) approximation to the expected avalanche size over an ensemble of realizations may be found by multiplying these probabilities:

$$S \approx \rho [1 - (1 - S_e)^{\lfloor \rho_0 N \rfloor}]. \quad (\text{A1})$$

When ρ_0 is independent of N (as assumed in most of our work), the exponent $\lfloor \rho_0 N \rfloor$ in this expression grows with N and the expected avalanche size approaches ρ as $N \rightarrow \infty$ —this justifies the approach in the main body of this paper. However, for single-seed activation we have $\rho_0 N = 1$ and the avalanche size is ρ with an adjustment factor S_e to give

$$S_{\text{single-seed}} \approx \rho S_e. \quad (\text{A2})$$

It remains only to calculate the size S_e of the extended vulnerable cluster. This was determined numerically in [13]; here we demonstrate how it may be calculated as an application of our general method. First, note (as was done in [13]) that the vulnerable cluster may be calculated as a site percolation problem with degree-dependent occupation probability $Q_k = F(1, k)$: using Eq. (9) in Eq. (4) we find

$$S_v = \sum_{k=0}^{\infty} p_k [1 - (1 - q_{\infty})^k] F(1, k), \quad (\text{A3})$$

where q_{∞} is found by iterating to steady state the relation from Eq. (1):

$$q_{n+1} = \sum_{k=1}^{\infty} \frac{k}{z} p_k [1 - (1 - q_n)^{k-1}] F(1, k). \quad (\text{A4})$$

These equations reproduce the results of Eqs. (3a) and (3b) of [13] for the size of the vulnerable cluster. We can also find the size S_e of the extended vulnerable cluster analytically (this was done only numerically in [13]). We simply alter the central-node relation (A3) so that all nodes with at least one child in the vulnerable cluster (each child having probability q_{∞} of being vulnerable) are included in the extended vulnerable cluster, thus giving

$$S_e = \sum_{k=0}^{\infty} p_k [1 - (1 - q_{\infty})^k]. \quad (\text{A5})$$

This equation is similar to Eq. (A3) but with the $F(1, k)$ term replaced with 1, while still using the same q_{∞} from Eq. (A4).

Figure 5 shows comparisons of the corrected mean avalanche size (A1) with numerical simulations for single-node and two-node initial seeds (i.e., $\rho_0 = 1/N$ and $\rho_0 = 2/N$, respectively) in Watts' model with response function (3) and uniform threshold level $R = 0.18$ [cf. Fig. 2(b) of [13]]. Despite the fact that the approximation (A1) is not expected to be valid near the percolation transition, we see rather good agreement with the numerical simulations. In the main body of this paper we take ρ_0 to be independent of N as $N \rightarrow \infty$ (numerics use seed fractions several magnitudes larger than

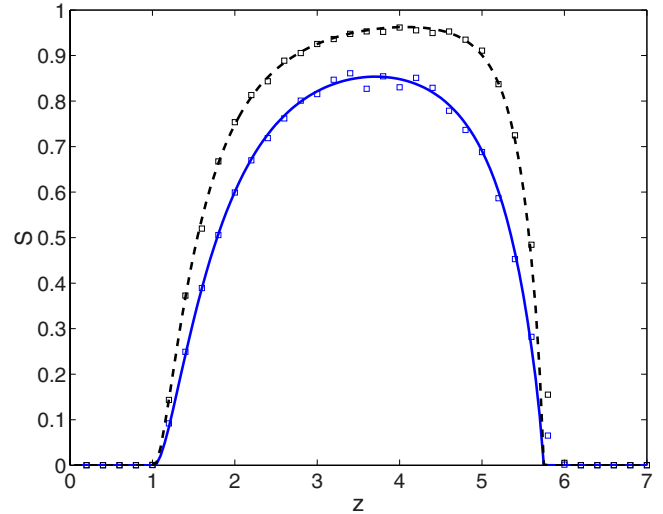


FIG. 5. (Color online) Average cascade size in Watts' model with uniform threshold $R = 0.18$ on Poisson random graphs with mean degree z . Solid line shows single-seed ($\rho_0 = 1/N$) and dashed line shows double-seed ($\rho_0 = 2/N$) predictions of Eqs. (A1) and (A5); symbols show results of numerical simulations on networks of size $N = 10^5$, averaged over 10^3 realizations.

$1/N$) so that ρ gives an excellent approximation to the measured mean cascade size, but note that (A1) should be employed in cases where $\rho_0 N$ is sufficiently small.

APPENDIX B: DERIVATION OF POISSON RANDOM GRAPH RESULT FROM [27]

The recent paper [27] examines cascades in the modified Watts model of Sec. II B 1 on modular networks with $d = 2$ communities. Each community has a Poisson degree distribution, and a pair of differential equations are derived to determine the size of network cascades. Here we demonstrate that the general asynchronous Eqs. (21)–(24) reduce to those in [27] under the assumptions used in that paper.

Exchanging the order of the k and m summations in the function $g^{(i)}$ defined in Eqs. (18) and (17), and dropping the time step index for convenience, we obtain

$$g^{(i)}(q) = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_{m=0}^{\infty} \sum_{k=m+1}^{\infty} \frac{k}{z^{(i)}} p_k^{(i)} \binom{k-1}{m} (\bar{q}^{(i)})^m \times (1 - \bar{q}^{(i)})^{k-1-m} F^{(i)}(m, k). \quad (\text{B1})$$

If $F^{(i)}(m, k)$ happens to be independent of k we may take it outside the k summation and then use the following result for the Poisson degree distribution $p_k^{(i)} = e^{-z^{(i)}} (z^{(i)})^k / k!$:

$$\sum_{k=m+1}^{\infty} \frac{k}{z^{(i)}} p_k^{(i)} \binom{k-1}{m} (\bar{q}^{(i)})^m (1 - \bar{q}^{(i)})^{k-1-m} = \frac{(\bar{q}^{(i)} z^{(i)})^m}{m!} e^{-\bar{q}^{(i)} z^{(i)}} \quad (\text{B2})$$

to obtain

$$g^{(i)}(q) = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_{m=0}^{\infty} F^{(i)}(m) \frac{x^m}{m!} e^{-x} \quad (\text{B3})$$

with $x = \bar{q}^{(i)} z^{(i)}$. For step function $F^{(i)}$ as in [27] the sum over m may be written in terms of the regularized incomplete gamma function.

Under the same assumptions on $F^{(i)}$ and $p_k^{(i)}$ it follows that the updating function $h^{(i)}$ for the density $\rho^{(i)}$ is identical to

the function $g^{(i)}$ above, and hence in the special case studied in [27] we have the relationship $q^{(i)} \equiv \rho^{(i)}$. This enables the solutions for $\rho^{(i)}(t)$ (in the asynchronous updating case) to be determined by solving only d nonlinear differential equations, whereas the general case of arbitrary degree distribution and/or k -dependent response functions requires the solution of $2d$ equations for the $q^{(i)}(t)$ and $\rho^{(i)}(t)$ functions, as given by Eqs. (21)–(24).

-
- [1] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
 [2] L. A. N. Amaral and J. M. Ottino, *Eur. Phys. J. B* **38**, 147 (2004).
 [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006).
 [4] E. J. Lee, K. I. Goh, B. Kahng, and D. Kim, *Phys. Rev. E* **71**, 056108 (2005).
 [5] Y. C. Lai, A. Motter, T. Nishikawa, K. Park, and L. Zhao, *Pramana, J. Phys.* **64**, 483 (2005).
 [6] P. Crucitti, V. Latora, and M. Marchiori, *Phys. Rev. E* **69**, 045104(R) (2004).
 [7] A. E. Motter and Y. C. Lai, *Phys. Rev. E* **66**, 065102(R) (2002).
 [8] Y. Moreno, R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Europhys. Lett.* **62**, 292 (2003).
 [9] P. Holme and B. J. Kim, *Phys. Rev. E* **65**, 066109 (2002).
 [10] K. I. Goh, D. S. Lee, B. Khang, and D. Kim, *Physica A* **346**, 93 (2005).
 [11] K. I. Goh, D. S. Lee, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **91**, 148701 (2003).
 [12] T. Antal, S. Redner, and V. Sood, *Phys. Rev. Lett.* **96**, 188104 (2006).
 [13] D. J. Watts, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766 (2002).
 [14] P. S. Dodds and D. J. Watts, *Phys. Rev. Lett.* **92**, 218701 (2004).
 [15] W. Duan, Z. Chen, Z. Liu, and W. Jin, *Phys. Rev. E* **72**, 026133 (2005).
 [16] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
 [17] D. Lopez-Pintado and D. J. Watts, preprint available from http://www.feem-web.it/ctn/papers/07_lopez-pintado.pdf
 [18] D. J. Watts and P. S. Dodds, *J. Consum. Res.* **34**, 441 (2007).
 [19] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Phys. Rev. Lett.* **96**, 040601 (2006).
 [20] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
 [21] D. Dhar, P. Shukla, and J. P. Sethna, *J. Phys. A* **30**, 5259 (1997).
 [22] M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).
 [23] B. Samuelsson and J. E. S. Socolar, *Phys. Rev. E* **74**, 036113 (2006).
 [24] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, *Phys. Rev. Lett.* **70**, 3347 (1993).
 [25] P. Shukla, *Int. J. Mod. Phys. B* **17**, 5583 (2003).
 [26] J. P. Gleeson and D. J. Cahalane, *Phys. Rev. E* **75**, 056103 (2007).
 [27] A. Galstyan and P. Cohen, *Phys. Rev. E* **75**, 036109 (2007).
 [28] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).
 [29] R. Cohen and K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
 [30] R. Cohen, D. ben-Avraham, and S. Havlin, *Phys. Rev. E* **66**, 036113 (2002).
 [31] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **73**, 056101 (2006).
 [32] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11150 (2007).
 [33] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
 [34] M. Molloy and B. Reed, *Random Structures Algorithms* **6**, 161 (1995).
 [35] M. Marder, *Phys. Rev. E* **75**, 066103 (2007).
 [36] J. P. Gleeson and D. J. Cahalane, *Proceedings of Noise and Stochastics in Complex Systems and Finance*, edited by J. Kerstész, S. Bornholdt, and R. N. Mantegna, Florence, Italy, *Proc. SPIE* 6601, 66010W (2007).
 [37] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
 [38] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
 [39] J. Hull and A. White, *J. Derivatives* **12**, 8 (2004).
 [40] M. A. Serrano and M. Boguñá, *Phys. Rev. E* **74**, 056114 (2006).