Seed size strongly affects cascades on random networks

James P. Gleeson[1] and Diarmuid J. Cahalane[2]
[1]*Applied Mathematics, University College Cork, Cork, Ireland*
[2]*MACSI, University College Cork, Cork, Ireland*
(Received 16 February 2007; published 3 May 2007)

The average avalanche size in the Watts model of threshold dynamics on random networks of arbitrary degree distribution is determined analytically. Existence criteria for global cascades are shown to depend sensitively on the size of the initial seed disturbance. The dependence of cascade size upon the mean degree z of the network is known to exhibit several transitions—these are typically continuous at low z and discontinuous at high z; here it is demonstrated that the low-z transition may in fact be discontinuous in certain parameter regimes. Connections between these results and the zero-temperature random-field Ising model on random graphs are discussed.

DOI: 10.1103/PhysRevE.75.056103 PACS number(s): 89.75.Hc

Dynamical models on networks have attracted a great deal of recent research interest; see the reviews [1–3] and references therein. Of particular interest are the effects of the topological structure of the network upon the time evolution of dynamical quantities. Under certain circumstances a change of state at a small number of network nodes may cause a cascade of changes over the whole network—such cascades may occur as, for example, overload failures [4–9], avalanches in sandpile models [10,11], evolution of species [12], or the spread of rumors and fads in social networks [13–15].

In this paper, we analyze a model of dynamics on random networks from the class of problems known as *binary decisions with externalities* [16]. Each node of an undirected network represents an agent, and may be in one of two states, called *active* and *inactive*. The network is chosen from an ensemble of graphs with a specified degree distribution p_k, i.e., the probability that a node has degree k is p_k, with $\Sigma p_k = 1$ [1,17]. Each agent also has a (frozen) random threshold r, chosen from a distribution with $F(r)$ denoting the probability that an agent has threshold less than r. Initially the network is *seeded* by activating a randomly chosen fraction ρ_0 of the N nodes. When updated, an inactive agent calculates the fraction of its neighbors who are currently active (m/k, where k is the degree of the node and m is the number of active neighbors), and it becomes active if this fraction exceeds his threshold r. Once active, an agent cannot become inactive; this will be referred to as the *permanently active property* (PAP). Isolated nodes (degree $k=0$) are never updated. It can be shown that (unlike some other sequential spin-flip models [18]) the steady state of the system does not depend on whether the updating of all agents is performed synchronously, or in a random asynchronous fashion.

Watts introduced the above model [13] to demonstrate the interaction between network topology and individual thresholds in the spreading of rumors and fads. He used percolation methods to determine whether a single activated initial agent can induce network-spanning cascades of activation. We focus here on calculating ρ, the average fraction of active nodes in the steady state, where the ensemble average is over realizations of networks and threshold values, and in the limit of infinite network size: $N \to \infty$. Watts' criterion for global cascades is recovered in the $\rho_0 \to 0$ limit (since $\rho_0 = 1/N$ for single-node initialization) but important differences are found when the seed size is finite. The quantity ρ measures the average avalanche size, and has recently been studied for dynamics on directed random graphs, with applications to random Boolean networks [19], and on small-world topologies [20]. The magnetization in the zero-temperature random-field Ising model (RFIM) [21,22] is also closely analogous to ρ; the agents' threshold values correspond to the local quenched random fields of the RFIM. There are, however, two important differences: (i) the RFIM does not exhibit the PAP, since spins may flip either up or down depending on their local field; (ii) the RFIM has not been studied on random graphs of arbitrary degree distribution. Analytical solutions for the RFIM have been developed in the mean-field case [21] and on Bethe lattices [22,23]. Our approach reduces to the latter in the case of regular random graphs, where each node has exactly z neighbors ($p_k = \delta_{kz}$ for integer z), provided that ρ_0 is zero.

Our main result is that the average final fraction ρ of active nodes is given by

$$\rho = \rho_0 + (1-\rho_0) \sum_{k=1}^{\infty} p_k \sum_{m=0}^{k} \binom{k}{m} q_\infty^m (1-q_\infty)^{k-m} F\left(\frac{m}{k}\right), \quad (1)$$

where q_∞ is the fixed point of the recursion relation

$$q_{n+1} = \rho_0 + (1-\rho_0) G(q_n) \quad \text{for } n = 0,1,2\ldots, \quad (2)$$

with $q_0 = \rho_0$, and the nonlinear function G is defined by

$$G(q) = \sum_{k=1}^{\infty} \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} q^m (1-q)^{k-1-m} F\left(\frac{m}{k}\right). \quad (3)$$

Here z is the mean degree of the network, $z = \Sigma k p_k$. We postpone the derivation of (1) to the end of the paper, and first examine some of its consequences.

Figure 1 shows ρ in the case of uniform thresholds (all nodes have identical threshold $r=R$) on Poisson (Erdős-Rényi) random graphs with degree distribution $p_k = e^{-z} z^k / k!$. In Fig. 1(a) ρ is color coded on the R,z parameter plane for seed fraction $\rho_0 = 10^{-2}$; Fig. 1(b) compares ρ at $R=0.18$ with numerical simulations for several ρ_0 values, as a function of

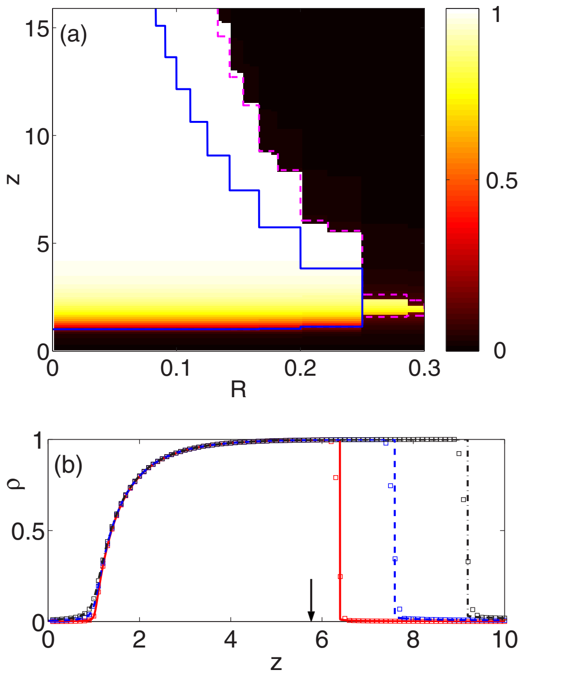

FIG. 1. (Color online) Average density ρ of active nodes in a Poisson random graph of mean degree z and uniform threshold value R. (a) Color-coded values of ρ from Eq. (1) on the R,z plane with seed fraction $\rho_0=10^{-2}$. Lines show approximations to the global cascade boundaries: the solid line encloses the region where (5) is satisfied; the dashed line represents the extended cascade boundary from Eq. (6). (b) Values of ρ at $R=0.18$ from Eq. (1) (lines) and numerical simulations (symbols), averaged over 100 realizations with $N=10^5$. Seed fractions are $\rho_0=10^{-3}$ (solid), 5×10^{-3} (dashed), and 10^{-2} (dot-dashed). The arrow marks the cascade boundary given by Eq. (5) in the limit of zero seed fraction.

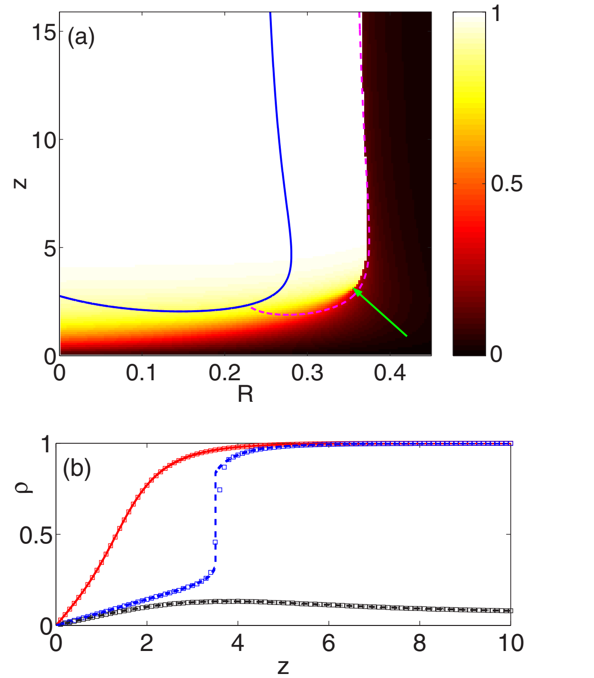

FIG. 2. (Color online) As Fig. 1, but threshold distribution is Gaussian with mean R and standard deviation 0.2, and seed fraction $\rho_0=0$. (a) Cascade boundaries as in Fig. 1; the arrow marks the critical point (R_c,z_c) described in the text; (b) theory and numerical simulation results ($N=10^5$, average over 100 realizations) at $R=0.2$ (solid), 0.362 (dashed), and 0.38 (dash-dotted).

z. The theory shows excellent agreement with simulations. Note the smooth transition from $\rho\approx0$ to $\rho\approx1$ near $z=1$, and the discontinuous transition back to $\rho\approx0$ at larger z values. The z location of the upper transition clearly depends sensitively on the initial activated fraction ρ_0.

A *global cascade* occurs with high probability when a small seed ρ_0 results in a large value of ρ. Writing $G(q)$ as $\Sigma_{\ell=0}^{\infty}C_\ell q^\ell$ with coefficients

$$C_\ell=\sum_{k=\ell+1}^{\infty}\sum_{n=0}^{\ell}\binom{k-1}{\ell}\binom{\ell}{n}(-1)^{\ell+n}\frac{k}{z}p_kF\left(\frac{n}{k}\right), \tag{4}$$

and linearizing Eq. (2) near $q=0$ gives a (first-order) condition for global cascades to occur: $(1-\rho_0)C_1>1$, since this guarantees that q_n increases with n, at least initially. This *cascade condition* may also be written as

$$\sum_{k=1}^{\infty}\frac{k(k-1)}{z}p_k\left[F\left(\frac{1}{k}\right)-F(0)\right]>\frac{1}{1-\rho_0}. \tag{5}$$

In the limit $\rho_0\to0$ and with $F(0)=0$, this reduces to the condition derived by Watts using percolation arguments [13]. On the R,z plane of Fig. 1(a) the condition (5) is satisfied inside the solid line; for $R=0.18$ in Fig. 1(b), (5) predicts cascade transitions at z values between 5.7 and 5.8 for both $\rho_0=10^{-3}$ and 10^{-2} (close to the z value marked by the arrow), but the simulations and full theory show that the respective transitions are actually near $z\approx6.4$ and 9.2. Condition (5) clearly does not accurately represent the effects of finite seed size [nor of nonzero $F(0)$; see below], because the function G is not well approximated by a straight line near the critical parameters for cascade transitions.

We seek an improved cascade condition by extending the series for $G(q)$ to order q^2. This approximation results in a quadratic equation for the fixed point q_∞ which we represent as $aq_\infty^2+bq_\infty+c=0$. Under the approximation of small q values, we interpret the existence of a positive root of this quadratic equation to imply $q_\infty\ll1$, and hence the impossibility of global cascades. Note that the first-order approximation (5) requires $b>0$ for global cascades. However, when the nonlinear behavior of G is important, it is still possible for global cascades to occur when b is negative, provided that the full solution of the quadratic equation precludes positive real roots. We therefore extend the cascade boundary to include regions where either $b>0$ (as in (5)) or $b^2-4ac<0$, the latter giving (to first order in ρ_0)

$$(C_1-1)^2-4C_0C_2+2\rho_0(C_1-C_1^2-2C_2+4C_0C_2)<0. \tag{6}$$

It is straightforward to plot this extended cascade condition: the dashed lines in Figs. 1(a) and 2(a) clearly give much improved approximations to the actual boundaries for global cascades.

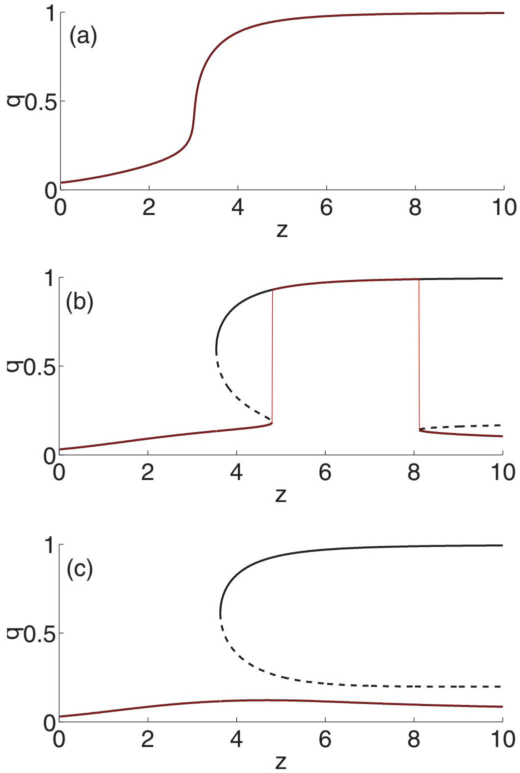

FIG. 3. (Color online) Bifurcation diagrams as described in text for dependence of q_∞ on z for $\sigma=0.2$ and $\rho_0=0$ at $R=$(a) 0.35, (b) 0.371, and (c) 0.375.

Figure 2(a) shows ρ on Poisson random graphs with thresholds drawn from a Gaussian distribution of mean R and standard deviation $\sigma=0.2$. Unlike the assumption in [13] that $F(0)=0$, a Gaussian distribution necessarily implies the presence of negative-valued thresholds among the population, so $F(0)>0$. Negative-threshold agents act as a natural seed, since they activate regardless of the states of their neighbors. The presence of these *innovators* [13] allows us to set $\rho_0=0$ in this case. The extended cascade condition (6) again gives a good approximation to the discontinuous ρ transition at high z values. Figure 2(b) focuses on the low-z transition and highlights the existence of a discontinuous transition in z for certain threshold distributions. This is qualitatively different from the previously-studied case [13] where only continuous low-z transitions were found.

Bifurcation analysis of Eq. (2) elucidates this result. In Fig. 3 we plot the roots of the fixed-point equations $G(q)-q=0$ (recall that $\rho_0=0$ here; extension to nonzero ρ_0 is straightforward) as functions of z, for different values of the mean threshold R. Thick solid and dashed lines denote stable and unstable fixed points respectively [24]. The PAP means the value of q_∞ achieved at a given z is that of the lowest stable branch above $q=\rho_0$. The occurrence of triple roots as R is increased causes the smooth low-z transition seen in Fig. 3(a) to become discontinuous [as shown by the thin solid line in Fig. 3(b)], as previously seen in the numerical simulations of Fig. 2(b). The discontinuous low-z transition occurs for $R>R_c$, where the critical coordinates (R_c,z_c) and the value $q=q_c$ where the triple root appears are found by numerical root finding for the system of three equations $q=\rho_0+(1-\rho_0)G(q)$, $(1-\rho_0)G'(q)=1$, and $G''(q)=0$. For $\sigma=0.2$ this yields $(R_c,z_c)=(0.3543,3.136)$; this point is marked with an arrow in Fig. 2(a). We remark that the discontinuous transition from $q_\infty\approx 1$ to $q_\infty\approx 0$ [which induces a similar transition in ρ through Eq. (1)] occurs due to a saddle-node bifurcation [24]. This behavior is quite generic, occurring for a wide variety of parameters, with the exception of the special case studied by Watts. For $\rho_0=0$ and $F(0)=0$ as in [13], the coefficient C_0 is zero and the fixed-point equation always has a root at $q=0$, with transcritical bifurcations on the $q=0$ line giving rise to the observed transitions. However, any nonzero seed size replaces the transcritical bifurcations with saddle-node bifurcations as described above. We have confirmed the accuracy of these results [and Eq. (1)] against numerical simulations on other configuration model network topologies [1], including power-law degree distributions (with exponential cutoff): $p_k \propto k^{-\tau}\exp(-k/\kappa)$ [17].

We turn now to the derivation of Eqs. (1)–(3). Our analytical approach is based on methods introduced by Dhar *et al.* to study the zero-temperature random-field Ising model on Bethe lattices [22]. The RFIM is a spin-based model of magnetic materials, and its zero-temperature limit has been extensively studied as a model for systems exhibiting hysteresis and Barkhausen noise [21]. A Bethe lattice of coordination number z (for integer z) is an infinite tree where every node has exactly z neighbors. Dhar *et al.* derive analytical results valid on Bethe lattices, but their numerical simulations show that the theory also applies very accurately to random graphs where every node has exactly z neighbors, provided that short-distance loops are rare. To analyze Watts' model we extend the approach of [22] in two ways. First, we consider treelike random graphs with arbitrary degree distributions, rather than the Bethe lattices of [22]. Second, we account for the PAP, which is the essential difference between Watts' update rule and standard RFIM dynamics. This difference between the update rules is crucial to our derivation of the ρ_0 dependence of the activated fraction ρ.

We begin by replacing the given random graph (with degree distribution p_k) by a tree structure. The top level of the tree is a single node with degree k, and this is connected to its k neighbors at the next lower level of the tree. Each of these nodes is in turn connected to k_i-1 neighbors at the next lower level, where k_i is the degree of node i. The degree distribution of the nodes in the tree is given by $\tilde{p}_k=(k/z)p_k$, which is the distribution for the number of nearest neighbors in a connected graph [1,25]. To find the final density ρ of active nodes, we label the levels of the tree from $n=0$ at the bottom, with the top node at an infinitely high level ($n\to\infty$). Define q_n as the conditional probability that a node on level n is active, conditioned on its parent (on level $n+1$) being inactive. Consider updating a node on level $n+1$, assuming that the nodes on all lower levels have already been updated. With probability $\tilde{p}_k$ the chosen node has k neighbors: one of these is its parent (on level $n+2$), and the remaining $k-1$ are its children (on level n). Since a fraction ρ_0 of nodes were initially set to be active, there is a probability ρ_0 that we have chosen one of these nodes. In this case the state of the node remains unchanged. On the other hand, with probability $(1-\rho_0)$ the node in question is inactive. In this case we must consider its neighbors. Each of the $k-1$ children is active with probability q_n. We also assume its parent is inactive. Thus the node has m active children (and therefore $k-1-m$ inactive children) with probability $\binom{k-1}{m}q_n^m(1-q_n)^{k-1-m}$. The probability that its threshold r_i is less than its fraction of active neighbors m/k is given by $F(m/k)$, and combining the independent probabilities yields Eq. (2) for q_{n+1}. The probability that the single node at the top of the tree is active is given by adding the probabilities for two independent cases: either it is already active as part of the originally activated fraction (with probability ρ_0), or it was initially inactive (with probability $1-\rho_0$). In the latter case, it will become active if sufficiently many of its k neighbors (all on the next lower level) are active. Noting that the top node has degree k with probability p_k, we conclude that the total probability of it being active is given by Eq. (1).

Although the theory is defined in terms of level-by-level updating on a tree, these results also apply to the random-graph Watts model, provided that (i) the network structure is locally treelike, (ii) the state of each node is altered at most once, and (iii) the steady state is independent of the order in which nodes are updated [23]. We note that condition (i) is true in configuration model networks, where the clustering coefficient scales as $1/N$ as $N\to\infty$ [1]. It does not, however, hold in small-world [20,26] and other real-world networks where clustering and loops play an important role. Condition (ii) is guaranteed by the PAP, and (iii) holds for this and related *unordered binary avalanche* models [19].

In summary, our main result is Eq. (1) for determining the expected fraction of active nodes (mean avalanche size) ρ due to the activation of a seed fraction ρ_0 of random nodes. The simple conditions for global cascades (5) and (6) follow; these substantially extend Watts' original result to include nonzero initial seed sizes. Values of ρ_0 as low as 0.1% have dramatic effects on the location of cascade transition points [see Fig. 1(b)]. Transitions in the dependence of ρ upon the mean network degree z are explained by the appearance of saddle-node bifurcation points. Of particular note is the fact that in certain parameter regimes the low-z transition may be discontinuous.

This work is funded by Science Foundation Ireland under Investigator Grant No. 02/IN.1/IM062 and the Mathematics Applications Consortium for Science and Industry in Ireland (MACSI), and by the Irish Research Council for Science, Engineering and Technology under the Embark Initiative, Grant No. RS/2005/16.

[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 *et al.*, 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 *et al.*, 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 *et al.*, Europhys. Lett. **62**, 292 (2003).
[9] P. Holme and B. J. Kim, Phys. Rev. E **65**, 066109 (2002).
[10] K. I. Goh *et al.*, 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. Q. Duan, Z. Chen, Z. Liu, and W. Jin, Phys. Rev. E **72**, 026133 (2005).
[16] T. C. Schelling, J. Conflict Resolut. **17**, 381 (1973).
[17] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[18] P. R. Eastham, R. A. Blythe, A. J. Bray, and M. A. Moore, Phys. Rev. B **74**, 020406(R) (2006).
[19] B. Samuelsson and J. E. S. Socolar, Phys. Rev. E **74**, 036113 (2006).
[20] D. Centola, V. M. Eguiluz, and M. W. Macy, Physica A **374**, 449 (2007).
[21] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, Phys. Rev. Lett. **70**, 3347 (1993).
[22] D. Dhar, P. Shukla, and J. P. Sethna, J. Phys. A **30**, 5259 (1997).
[23] P. Shukla, Int. J. Mod. Phys. B **17**, 5583 (2003).
[24] S. H. Strogatz *Nonlinear Dynamics and Chaos* (Perseus Books, Cambridge, MA, 1994).
[25] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[26] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).