Mean size of avalanches on directed random networks with arbitrary degree distributions

James P. Gleeson
Department of Mathematics & Statistics, University of Limerick, Ireland

(Received 1 October 2007; revised manuscript received 23 March 2008; published 19 May 2008)

The mean size of unordered binary avalanches on infinite directed random networks may be determined using the damage propagation function introduced by [B. Samuelsson and J. E. S. Socolar, Phys. Rev. E **74**, 036113 (2006)]. The derivation of Samuelsson and Socolar explicitly assumes a Poisson distribution of out-degrees. It is shown here that the damage propagation function method may be used whenever the in-degree and out-degree of network nodes are independently distributed; in particular, it is not necessary that the out-degree distribution be Poisson. The general case of correlated in- and out-degrees is discussed and numerical simulations (on large finite networks) are compared with the theoretical predictions (for infinite networks).

DOI: 10.1103/PhysRevE.77.057101 PACS number(s): 89.75.Da, 02.50.Ey, 02.10.Ox, 05.50.+q

Unordered binary avalanches (UBAs) on directed networks are defined in [1] and used to infer several interesting results for the dynamics of cascades on random Boolean networks. Percolation transitions and avalanche size distributions are characterized by a single function termed the *damage propagation function*. The analysis of [1] is based on the assumption of Poisson-distributed out-degrees and thus explicitly excludes, for example, regular random graphs. The purpose of the present note is to show that for large networks the mean avalanche size, and hence percolation transitions, may be found using the damage propagation function even when the out-degree distribution is not Poisson. Indeed this single function is shown to be sufficient to determine the mean avalanche size whenever the in-degree and out-degree of nodes are independently chosen from arbitrary distributions (including the case of random regular graphs). However, if the in-degree and out-degree of nodes are correlated then a single function is no longer sufficient to determine the mean avalanche size.

The mean avalanche size in a network of N binary-valued nodes is calculated by determining the number n of nodes with value 1 ("damaged nodes") in each realization once the avalanche has run to completion, and then averaging the fraction n/N over the ensemble of random networks [1]:

$$\phi = \lim_{N\to\infty} \langle n/N \rangle. \tag{1}$$

Our analytical results are derived under the infinite-network assumption $N\to\infty$, but show good agreement with numerical simulations for large but finite networks (see Figs. 1 and 3 for examples). As noted in [1], the value of ϕ acts as an order parameter for certain percolation transitions: the sparse percolation (SP) transition occurs when ϕ changes from zero to a nonzero value, while the exhaustive percolation (EP) transition is found when ϕ reaches 1. The quantity ϕ also plays an important role in threshold-decision models [2,3]—in these applications (on undirected networks) ϕ gives the expected fraction of damaged nodes after cascades have run to completion. Of particular interest in all cases are UBAs that are initiated by damage at a nonzero fraction ρ of *seed nodes*.

The damage propagation function $g(x)$ is defined as the probability that a random node will be damaged, given that each input is damaged with probability x. Given the in-degree distribution $P_{\mathrm{in}}(j)$ of the network (in the $N\to\infty$ limit) and the probability $P_0(j,i)$ that a node of in-degree j becomes damaged if i of its inputs are damaged, the damage propagation function may be written as

$$g(x) = \rho + (1-\rho)\sum_j P_{\mathrm{in}}(j)G(j,x), \tag{2}$$

where we introduce the notation G for the frequently occurring function

$$G(j,x) = \sum_{i=0}^{j} \binom{j}{i} x^i (1-x)^{j-i} P_0(j,i). \tag{3}$$

Note that we have slightly generalized the definition of [1] to include the nonzero probability ρ of a node being damaged initially: Eq. (A1) of [1] is recovered by setting $\rho=0$ in Eq. (2) above. We also assume here that the graph is connected;

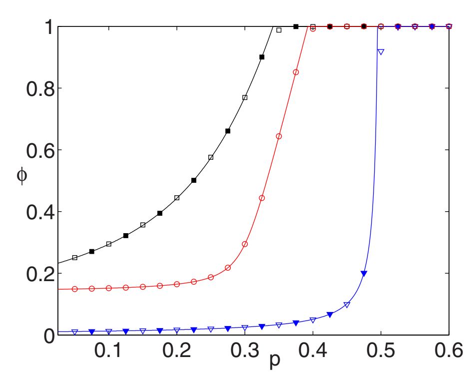

FIG. 1. (Color online) The order parameter ϕ for the exhaustive percolation example described in the text, as a function of the percolation parameter p. Curves show the predictions of Eqs. (4)–(6); symbols show the average fraction of damaged nodes over 100 realizations of random networks of 10^5 nodes. Open symbols are for Poisson out-degrees and closed symbols are for regular out-degrees (i.e., each node having exactly two outputs); in all cases the mean degree is $z=2$ and initial damage fraction is $\rho=0.01$. Left curve and squares: Poisson-distributed in-degrees. Right curve and triangles: regular in-degrees (i.e., each node has exactly two inputs). Middle curve and circles: Poisson-distributed in-degrees, with out-degree of each node set equal to its in-degree.

the case of disjoint clusters within the network will be considered further below.

Consider the case where all nodes are updated synchronously at time steps $m=1,2,\ldots$ (the order independence of the UBA implies that the same steady-state results are found for synchronous or asynchronous updating). The ensemble-averaged fraction of damaged nodes at time step m is denoted by ϕ_m and its steady state (as $m\to\infty$) gives the mean avalanche size ϕ defined in Eq. (1).

Defining x_m as the probability at time step m that a randomly chosen input to a randomly chosen node has value 1, we obtain the expression

$$\phi_{m+1}=\rho+(1-\rho)\sum_j P_{\rm in}(j)G(j,x_m)=g(x_m), \tag{4}$$

with $\phi_0=\rho$ [4]. In order to iterate this equation to determine the expected fraction of damaged nodes as $m\to\infty$, we require an updating equation for x_m. We consider the general case where the network has a joint probability distribution for in- and out-degrees [5]: $P(j,k)$ being the probability that a random node has in-degree j and out-degree k. Then the out-degree distribution of nodes which input to a randomly selected node is proportional to $kP(j,k)$, reflecting the fact that nodes with high out-degrees k are more likely to occur as inputs to our initially chosen node [5]. The appropriately normalized joint in- and out-degree distribution for nodes which input to a selected node is thus $k/zP(j,k)$, [where $z=\Sigma_{j,k}jP(j,k)=\Sigma_{j,k}kP(j,k)$ is the mean degree of the network] and x_m is updated as

$$x_{m+1}=\rho+(1-\rho)\sum_{k,j}\frac{k}{z}P(j,k)G(j,x_m), \tag{5}$$

with $x_0=\rho$. Iteration of Eqs. (4) and (5) enables us to calculate the mean avalanche size for networks with arbitrary joint distributions $P(j,k)$ of in- and out-degrees.

Considerable simplification of Eqs. (4) and (5) occurs if the joint distribution factorizes so that $P(j,k)=P_{\rm in}(j)P_{\rm out}(k)$. This is the case, for example, if the in-degree and out-degree of each node are independent random numbers. It also applies if all in-degrees or all out-degrees are equal, as is the case for regular random graphs. For such special cases, the summation over k in Eq. (5) may be performed to give an iteration for x_m in terms of the damage propagation function:

$$x_{m+1}=g(x_m). \tag{6}$$

Comparing with Eq. (4) we see that in this case the quantities ϕ_m and x_m are identical at all time steps, and so a single iteration equation [equivalently, the single function $g(x)$] suffices to determine the order parameter ϕ. Note that provided $P(j,k)$ factorizes, no assumption on the form of $P_{\rm out}(k)$ is needed, and in particular it is not necessary that $P_{\rm out}$ be Poisson as in [1]. However, for general networks with nonfactorizing $P(j,k)$ a single function is no longer sufficient to determine the mean avalanche size (and hence the percolation transitions), as Eq. (5) cannot then be expressed in terms of the damage propagation function (2).

To demonstrate the accuracy of these results we consider the model of exhaustive percolation (EP) proposed in [1]. In this adjusted bond-percolation problem all zero-input nodes are initially damaged, as are a fraction ρ of the remaining nodes. Also, any node whose inputs are all damaged becomes damaged itself. Each directed link in the network is set to transmit damage with probability p, as in standard bond percolation. The function $P_0(j,i)$ for this model is $1-(1-p)^i$ for $i<j$, with $P_0(j,j)=1$ and the sums over both j and k in the various iteration equations may be easily obtained in closed form [1]. Figure 1 demonstrates that results of numerical simulations for $N=10^5$ nodes (symbols) match the theoretical predictions for $N\to\infty$ (curves) of Eqs. (4)–(6) very well. Recall that our claim is that the mean avalanche size does not depend on the out-degree distribution, provided that $P(j,k)$ factorizes. This is supported by the fact that for the cases with uncorrelated in- and out-degrees both open symbols (Poisson out-degrees with $z=2$) and closed symbols (every out-degree equal to $z=2$) fall on the same theoretical curve—the case of Poisson in-degrees is shown with squares (leftmost curve), while the regular in-degree case $[P_{\rm in}(j)=\delta_{jz}]$ is shown with triangles (rightmost curve).

All networks are generated using the "configuration model" described in [6]: for each node a pair of (possibly correlated) random integers (j,k) are selected and the node is endowed with j "in-stubs" and k "out-stubs." Having generated the stubs for all nodes, each in-stub is randomly connected to an out-stub to form a link of the network. When all the stubs are connected [7] the resulting network is a realization of an ensemble characterized by the joint distribution $P(j,k)$ (which may, for example, have power-law tails). Self-links and short loops occur with probabilities on the order of $1/N$ and so may be neglected in the $N\to\infty$ limit [6].

To generate an example of a strongly correlated network with Poisson in- and out-degree distributions we constrain the out-degree of every node to equal its in-degree. The accuracy of our extended result for this case of $P(j,k)=\delta_{jk}P_{\rm in}(j)$ is shown in Fig. 1 by the match of circles (numerical simulations) to the theory of Eqs. (4) and (5) (middle curve). Note that the joint distribution $P(j,k)$ is *not* equal to $P_{\rm in}(j)P_{\rm out}(k)$ in this case and so the single-function simplification of Eq. (6) is not applicable here.

In Fig. 2 we compare the value of the order parameter ϕ on the (z,p) parameter plane for the same EP problem and with no initial damage: $\rho=0$. The white-colored region corresponds to $\phi=1$ and hence to exhaustive percolation; this allows a direct comparison with Fig. 3 of [1]. Both in- and out-degree distributions are Poisson for both panels of Fig. 2, but for (a) the in- and out-degrees are uncorrelated (as in [1]), while in (b) the in-degree of each node equals its out-degree. The effect of the correlations on the EP region is clearly dramatic.

This approach may also be applied to the sparse percolation (SP) problem considered briefly in [1]. The SP transition occurs for parameter values where an infinitesimally small seed fraction ρ is sufficient to generate nonzero average avalanche sizes. For this standard bond percolation problem, Eqs. (4) and (5) are applicable [cf. Eqs. (83)–(86) of [8]] with $P_0(j,i)=1-(1-p)^i$ and taking the limit of zero seed fraction, $\rho\to 0$. As noted in [1], the SP transition point may be located for the uncorrelated case by examining the slope of $g(x)$ at $x=0$; a general criterion for correlated in- and

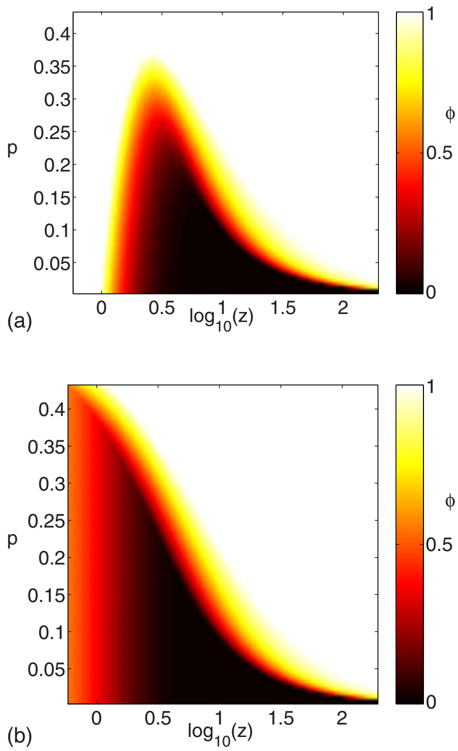

FIG. 2. (Color online) Order parameter ϕ on the (z,p) plane for networks with Poisson-distributed in- and out-degrees, with $z=2$ and $\rho=0$. (a) Uncorrelated in- and out-degrees, as in Fig. 3 of [1]; (b) correlated case, with in- and out-degree of each node being equal.

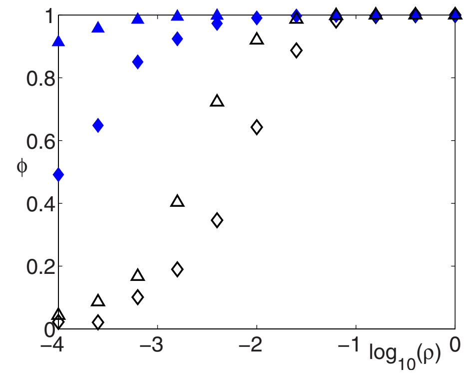

FIG. 3. (Color online) Order parameter ϕ for the SP case with full bond occupation ($p=1$) on networks with one input per node, as a function of seed fraction size ρ and averaged over 100 network realizations. Open symbols denote Poisson out-degrees, and closed symbols denote regular out-degrees (exactly one output per node). Convergence to the theoretical result $\phi\equiv 1$ for infinite networks is indicated by increasing the network size from $N=10^4$ (diamonds) to $N=10^5$ (triangles). Note that the smallest seed fraction shown, $\rho=10^{-4}$, corresponds to a single node being seeded in the networks of size $N=10^4$.

out-degrees follows from applying the same argument to the slope of the right-hand side of Eq. (5) at $x=0$. It easily follows that the critical value of p for bond percolation is given by [6,9]

$$p_c = \frac{z}{\sum_{j,k} jkP(j,k)}. \tag{7}$$

In the uncorrelated case this reduces to $p_c=1/z$, as found in Eq. (23) of [1]. It is not necessary for P_{out} to be Poisson and in particular this result holds for the regular random graphs which are specifically excluded in the discussion of Eq. (23) in [1].

Strictly speaking, our derivation of Eq. (4) implies that the value of ϕ obtained is the fraction of damaged nodes within the connected component(s) of the network which are accessible from the seed node(s). When $z\geq 2$ as in Figs. 1 and 2, the giant connected component essentially includes the whole network and so ϕ admits the interpretation of overall network damage fraction as above. However, in cases where the network is composed of disconnected clusters only those clusters which are successfully seeded by the initially damaged fraction ρ can attain the average damage level ϕ predicted by Eq. (4). The effect of this is particularly marked for the sparse percolation problem where single nodes are seeded to give $\rho=1/N$, with the limit $\rho\to 0$ as $N\to\infty$. We therefore expect to see deviations from the predictions of Eq. (4) for SP when ρ is vanishingly small and the network graph is poorly connected [but we note the transition point (7) is not affected by this].

Figure 3 shows numerical results for the average fraction of damaged nodes in the SP problem with $p=1$ on networks with exactly one input per node. This $z=1$ case is precisely at the percolation transition point (7) and so gives a clear demonstration of the differences between the Poisson out-degree (open symbols) and regular out-degree (closed symbols) cases on finite-sized networks with $N=10^4$ (diamonds) and $N=10^5$ (triangles) nodes. These results should be compared with the theoretical prediction for the $N\to\infty$ limit: Eq. (5) gives

$$x_{m+1} = \rho + (1-\rho)x_m \tag{8}$$

for this case which implies $x_{m+1}>x_m$ for any nonzero seed fraction ρ. Consequently we obtain $x_\infty=1$, with order parameter $\phi=1$. For any fixed (i.e., N-independent) value of ρ we see convergence toward this theoretical value as N is increased. However, because the distribution of connected in-component sizes depends on the out-degree distribution of the network [see, for example, Eq. (85) of [8]] we observe a strong dependence on the out-degree distribution when the seed fraction ρ is not large enough to activate all the disconnected clusters in each network. We conclude that Eqs. (4) and (5) correctly predict expected damaged fractions on the whole network if either: (i) the network is sufficiently well connected, or (ii) the fraction ρ of initially damaged nodes is sufficiently large to ensure all disconnected clusters are seeded.

In summary, we have shown that if the joint in- and out-degree distribution $P(j,k)$ may be factored as $P_{\text{in}}(j)P_{\text{out}}(k)$ then the single function $g(x)$ defined in [1] is sufficient to determine the average avalanche size in infinite connected networks via the iterated mapping (6). This result allows the determination of percolation transitions but does not require that P_{out} be Poisson, and so generalizes the work of [1]. In the general case where the in- and out-degrees are correlated we have demonstrated that the computation of the average avalanche size requires two iterating functions [Eqs. (4) and (5)] rather than the single damage propagation function of [1]. These results on the average avalanche size are valid for networks in the $N \to \infty$ limit with arbitrary degree distributions [4] generated by the configuration model described above, provided (as demonstrated in a critical case in Fig. 3) that the network is well connected or the number of seed nodes ($\approx \rho N$) is sufficiently large. Our analysis assumes the infinite-network size limit, but results of numerical simulations for networks of finite (but large) size show excellent agreement with the theory.

While we have shown that the mean size of avalanche on networks in the infinite-size limit may be calculated under more general conditions than those considered in [1], much work remains to be done. We note from our numerical simulations that even when $P(j,k)$ factorizes the full probability distribution function (pdf) of avalanche sizes shows a dependence on the type of out-degree distribution, despite the fact the mean avalanche size (i.e., first moment of the pdf) is independent of the out-degree distribution in such cases. Further work is therefore required to extend the theory for the Poisson out-degree case examined in [1] to find the full distribution of avalanche sizes in networks with arbitrary degree distributions.

This work was funded by Science Foundation Ireland under programs 06/IN.1/I366, MACSI, and 05/RFP/MAT0016.

[1] B. Samuelsson and J. E. S. Socolar, Phys. Rev. E **74**, 036113 (2006).
[2] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[3] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).
[4] We assume that the probability of an input being 1 is independent for distinct inputs to a given node. This is true for large networks with well-behaved degree distributions, but excludes networks with hubs which output to a substantial fraction of the nodes.
[5] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[7] Realizations where the total number of in-stubs differs from the total number of out-stubs may be dealt with by re-drawing the random numbers until the totals are equal [6].
[8] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, J. Theor. Biol. **240**, 400 (2006).
[9] N. Schwartz, R. Cohen, D. ben-Avraham, A. L. Barabasi, and S. Havlin, Phys. Rev. E **66**, 015104(R) (2002).