Weighted Percolation on Directed Networks

Juan G. Restrepo,[1,*] Edward Ott,[2] and Brian R. Hunt[2]
[1]Northeastern University, Boston, Massachusetts 02115, USA
[2]University of Maryland, College Park, Maryland 20742, USA
(Received 4 April 2007; published 4 February 2008)

We present and numerically test an analysis of the percolation transition for general node removal strategies valid for locally treelike directed networks. On the basis of heuristic arguments we predict that, if the probability of removing node i is p_i, the network disintegrates if p_i is such that the largest eigenvalue of the matrix with entries $A_{ij}(1-p_i)$ is less than 1, where A is the adjacency matrix of the network. The knowledge or applicability of a Markov network model is not required by our theory, thus making it applicable to situations not covered by previous works.

DOI: 10.1103/PhysRevLett.100.058701 PACS numbers: 89.75.Hc, 05.70.Ln, 87.23.Ge

There has been much recent interest in the structure and function of complex networks [1]. One aspect that has received considerable attention is the resilience of networks to the removal of some of their nodes [2–11]. This problem is related, for example, to the robustness of transportation and information networks to disturbances like random failures or targeted attacks, or to the resistance of biological networks to the action of drugs. Another related problem is determining the threshold for epidemic spreading [12]. An important objective is to determine what characteristics of the structure of the network determine the proportion of nodes that can be removed before the network disintegrates.

A model often considered is one in which nodes are removed from a network of size N with a uniform probability p. In the large N limit, for probabilities less than a critical value p_c, there is a connected component of the network of size of order N (the *giant component*), and for values of p larger than p_c, there is no connected component of size of order N [1,3–8,10]. The critical probability p_c at which this *percolation* transition occurs has been the subject of several theoretical works.

In what follows we define the in and out degrees of a node i by $d_i^{\leftrightarrow} = \sum_{j=1}^{N} A_{ij}$ and $d_i = \sum_{j=1}^{N} A_{ji}$. Here A_{ij} is the network adjacency matrix; $A_{ij} = 1$ if there is a directed link from i to j and 0 otherwise. If $A = A^T$ the network is said to be undirected and $d_i^{\leftrightarrow} = d_i = d_i$. For undirected, degree uncorrelated networks (the number of connections per node for neighboring nodes is not correlated), Cohen *et al.* have shown [4] that the critical probability is approximately given by $(1-p_c)[(\langle d^2\rangle/\langle d\rangle) - 1] = 1$, where $\langle\cdot\rangle$ denotes an average over nodes. Reference [8] treats the case of undirected networks with correlations for *degree Markovian* networks, i.e., networks in which all nontrivial correlations are captured by the probability $P(d'|d)$ that a randomly chosen link from a node with degree d is connected to a node with degree d'. Other works on undirected networks have extended the Markovian approach to include the effect of clustering (e.g., often present in social networks); in particular, Ref. [11] presents an analysis for the case of weak clustering.

Reference [9] first studied the percolation transition in directed degree Markovian networks. The types of components studied are a *strongly connected component* (SCC), defined as a set of nodes such that every node in the SCC is reachable from any other node in the SCC by a directed path, its associated *in-component* (IN), defined as the set of nodes from which the SCC is reachable by a directed path, and *out-component* (OUT), defined as the set of nodes reachable from the SCC by a directed path. (There might be several such components.) Of interest is the largest strongly connected component which, if its size is of order N, is called the *giant strongly connected component*, GSCC. The out and in components of the GSCC are denoted GOUT and GIN.

It was found in Ref. [10] that as the probability of node removal p increases, GSCC, GOUT, and GIN disappear at the same critical value p_c. This value is determined by the largest eigenvalue of a matrix expressed in terms of $P_o(\mathbf{y}'|\mathbf{y})$ and $P_b(\mathbf{y}'|\mathbf{y})$ where $\mathbf{y} = (d_p, d_p^{\leftrightarrow}, d_p^{\leftrightarrow})$, and d_p, $d_p^{\leftrightarrow}$ and $d_p^{\leftrightarrow}$ are the number of incoming, outgoing and bidirectional edges for a given node. Here $P_o(\mathbf{y}'|\mathbf{y})$ and $P_b(\mathbf{y}'|\mathbf{y})$ are the probabilities of reaching a node of degree $\mathbf{y}'$ from a node of degree $\mathbf{y}$ by following an outgoing and a bidirectional edge, respectively.

One of our aims in this Letter is to remove the need for the applicability and knowledge of a Markov network model. In order to do so, we will focus on a class of directed networks that are locally treelike in the sense that they have few short loops [13]. More precisely, we assume that for each node i and not too large L, the number of *different* nodes reachable by paths of length L or less starting at node i is close to the *total* number of paths of length L or less starting from node i. In particular ($L = 2$) we assume that bidirectional edges are negligible in number. Under this assumption, $\mathbf{y} = (d, d^{\leftrightarrow}, 0)$, and the matrix in Ref. [10] whose eigenvalue determines p_c reduces to

 © 2008 The American Physical Society

$$C_{\mathbf{z}\mathbf{z}'} = (d^{\blacktriangleright})' P(\mathbf{z}'|\mathbf{z}), \tag{1}$$

where $\mathbf{z} = (d, d^{\blacktriangleright})$. We note that our locally treelike condition for directed networks is analogous to assuming negligible clustering.

In many situations the node removal probability is not a constant. For example, airports might have different security measures, or differ in their vulnerability to an attack or weather related shutdown due to their geographical location. Also, we have noted recently [14] that a measure of the dynamical importance of node i is proportional to $v_i u_i$, where u and v are the right and left eigenvectors corresponding to the largest eigenvalue λ of the adjacency matrix of the network, A: $\mathrm{A} = \lambda u, vi$

m, the number of paths of length m grows like λ^m, we associate to the previous equation the eigenvalue problem $\lambda_M \psi_{\mathbf{z}} = d \sum_{\mathbf{z}'} P(\mathbf{z}'|\mathbf{z}) \psi_{\mathbf{z}'}$, where λ_M is the *Markovian approximation* to λ. The previous result agrees with Eq. (1) [the matrices $d P(\mathbf{z}'|\mathbf{z})$ and $(d)' P(\mathbf{z}'|\mathbf{z})$ have the same spectrum]. We note that, in the absence of degree-degree correlations, we have $P(\mathbf{z}'|\mathbf{z}) = d P(\mathbf{z}')/\langle d \rangle$, which yields the *mean field approximation* for the eigenvalue, $\lambda_{mf} = \langle d \rangle / \langle d \rangle$. This agrees with the results in Ref. [10], where the effect of bidirectional edges is considered, and, for uncorrelated networks, a formula interpolating between λ_{mf} (when bidirectional edges are rare) and the undirected result $(\langle d^2 \rangle / \langle d \rangle) - 1$ is found. This supports our claim that, if short closed paths are rare, then λ should be a good approximation to the threshold.

We now illustrate our theory with two numerical examples and one real network. Example 1 illustrates the flexibility of our approach to address various weighted percolation node removal strategies, while example 2 illustrates the point that our approach does not require the knowledge or applicability of a Markov network model. We note that the networks in consideration are sparse, which allows us to use efficient techniques to compute the largest eigenvalue.

Example 1.—For simplicity, we consider uncorrelated random networks with degree distributions $P(d, d)$ in which d and d are independent and have the same distribution $P(d)$, that is, $P(d, d) = P(d) P(d)$. We use a generalization of the method in Ref. [16] in order to generate networks with a power law degree distribution, $P(d) \propto d^{-\gamma}$. We choose the sequence of expected degrees $d_i = c(i + i_0 - 1)^{-1/(\gamma-1)}$ for the in-degrees, and a random permutation of this sequence for the out-degrees, where $i = 1, \ldots, N$, and c and i_0 are chosen to obtain a desired maximum and average degree. Then, the adjacency matrix is constructed by setting $A_{ij} = 1$ for $i \neq j$ with probability $d_i d_j / (N \langle d \rangle)$ and zero otherwise ($A_{ii} = 0$). The ensemble expected value of the resulting network degree distribution is given by $P(d, d)$. (Note that we assume $d_i d_j < N \langle d \rangle$.) In Fig. 1(a) we show, for a $N = 2000$ scale free network with exponent $\gamma = 2.5$ and $\langle d \rangle = 3$, the size of GIN as a function of the number of removed

degree distribution is given71.491 Tf3.4089 0.2e24 0 TD[(,)-264.9(to)6.4(ved)]TJ8 average degree. Then, the adjacency

j

Example 2.—We start with a network generated as in example 1 with $N = 10^3$, $\gamma = 2.5$, and $\langle d \rangle = 3$. Then, we first specify a division of the nodes in the network into two groups of the same size, X and Y. We define a measure of the degree-degree correlations $\rho = \langle d_i\, d_j \rangle_e / \langle d_i \rangle_e^2$, where $\langle \ldots \rangle_e$ denotes an average over edges, $\langle Q_{ij} \rangle_e \equiv \sum_{i,j} A_{ij} Q_{ij} / \sum_{i,j} A_{ij}$. The following (an adaptation of the method in Ref. [21]) is repeated until the network has the desired amount of degree-degree correlations as evidenced in the value of ρ: Two edges are chosen at random, say connecting node i to node j and node n to node m. If i, j, n, m are all in X and $(d_n d_m + d_i d_j - d_n d_j - d_i d_m) < 0$, the edges are replaced with two edges connecting node i to node m and node n to node j. Otherwise the edges are unchanged. By repeating this process for several steps one creates two subnetworks, X and Y, with different degree-degree correlations (here, 2.29 and 0.98, respectively). Starting from such a network, we successively remove a randomly chosen node and compute the size of the GIN relative to its initial size. In Fig. 2 we plot this normalized size of the GIN as a function of the fraction of remaining nodes $(1 - p)$ for ten realizations of the node removal sequence. Although the transition points of individual realizations have some spread, the arrow (predicted from the eigenvalue) gives a good approximation of their mean (see inset). Similar results for both examples 1 and 2 were obtained for tests using other values of the network parameters γ, $\langle d \rangle$, and N.

We now discuss the advantages and disadvantages of the eigenvalue approach when compared to the Markov approximation. As opposed to the Markov approximation, the eigenvalue approximation allows the easy treatment of general node removal strategies (''weighted percolation''). Furthermore, it does not require the assumption that the node correlations depend only on their degree and are only to nearest neighbors. In addition, the construction of the matrix $P(\mathbf{z}'|\mathbf{z})$ and the determination of its largest eigenvalue is in some cases harder than the direct determination of the largest eigenvalue of the adjacency matrix A. On the other hand, in many cases the adjacency matrix of the network is not known, and local sampling methods from which an approximation to the matrix $P(\mathbf{z}'|\mathbf{z})$ can be constructed must be used. Additionally, the eigenvalue approach is valid only when the network has locally treelike structure. As such, our method should be viewed as complementary to the Markov approach.

This work was supported by ONR (Physics), by the NSF (No. PHY 0456240 and No. DMS 0104-087), and by USAFOSR No. FA95500410319.

*juanga@neu.edu

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003); A.-L. Barabási and R. Albert, Rev. Mod. Phys. **74**, 47 (2002); S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006).
[2] R. Albert *et al.*, Nature (London) **406**, 378 (2000).
[3] D. S. Callaway *et al.*, Phys. Rev. Lett. **85**, 5468 (2000).
[4] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).
[5] R. Cohen *et al.*, Phys. Rev. Lett. **91**, 247901 (2003).
[6] L. K. Gallos *et al.*, Phys. Rev. Lett. **94**, 188701 (2005).
[7] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[8] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).
[9] N. Schwartz *et al.*, Phys. Rev. E **66**, 015104(R) (2002).
[10] M. Boguñá 74, 015554(1.1616.83199s.1561 TD [([2])5).