PHYSICAL REVIEW E **66**, 015104(R) (2002)

Percolation in directed scale-free networks

N. Schwartz,[1] R. Cohen,[1] D. ben-Avraham,[2] A.-L. Barabási,[3] and S. Havlin[1]
[1]*Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan, Israel*
[2]*Department of Physics, Clarkson University, Potsdam, New York 13699-5820*
[3]*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556*
(Received 1 May 2002; published 26 July 2002)

Many complex networks in nature have directed links, a property that affects the network's navigability and large-scale topology. Here we study the percolation properties of such directed scale-free networks with correlated *in* and *out* degree distributions. We derive a phase diagram that indicates the existence of three regimes, determined by the values of the degree exponents. In the first regime we regain the known directed percolation mean field exponents. In contrast, the second and third regimes are characterized by anomalous exponents, which we calculate analytically. In the third regime the network is resilient to random dilution, i.e., the percolation threshold is $p_c \to 1$.

DOI: 10.1103/PhysRevE.66.015104 PACS number(s): 02.50.Cw, 05.50.+q, 64.60.Ak

Recently the topological properties of large complex networks such as the Internet, the World Wide Web (WWW), an electric power grid, and cellular and social networks have drawn considerable attention [1,2]. Some of these networks are directed, for example, in social and economical networks if node *A* gains information or acquires physical goods from node *B*, it does not necessarily mean that node *B* gets similar input from node *A*. Likewise, most metabolic reactions are one directional, thus changes in the concentration of molecule *A* affect the concentration of its product *B*, but the reverse is not true. Despite the directedness of many real networks, the modeling literature, with few notable exceptions [3–5], has focused mainly on undirected networks.

An important property of directed networks can be captured by studying their degree distribution, $P(j,k)$, or the probability that an arbitrary node has *j* incoming and *k* outgoing edges. Many naturally occurring directed networks, such as the WWW, metabolic networks, citation networks, etc., exhibit a power-law, or *scale-free* degree distribution for the incoming or outgoing links:

$$P_{in(out)}(l) = c l^{-\lambda_{in(out)}}, \quad l \geq m, \tag{1}$$

where *m* is the minimal connectivity (usually taken to be $m=1$), *c* is a normalization factor, and $\lambda_{in(out)}$ are the in (out) degree exponents characterizing the network [6,7]. An important property of scale-free networks is their robustness to random failures, coupled with an increased vulnerability to attacks [8–12]. Recently it has been recognized that this feature can be addressed analytically in quantitative terms [9–11] by combining graph theoretical concepts with ideas from percolation theory [13–15]. Yet, while the percolative properties of undirected networks are much studied, little is known about the effect of node failure in directed networks. As many important networks are directed, it is important to fully understand the implications to their stability. Here we show that directedness has a strong impact on the percolation properties of complex networks and we draw a detailed phase diagram.

The structure of a directed graph has been characterized in [3,4], and in the context of the WWW in [7]. In general, a directed graph consists of a giant weakly connected component (GWCC) and several finite components. In the GWCC every site is reachable from every other, provided that the links are treated as bidirectional. The GWCC is further divided into a giant strongly connected component (GSCC), consisting of all sites reachable from each other following directed links. All the sites reachable from the GSCC are referred to as the giant OUT component, and the sites from which the GSCC is reachable are referred to as the giant IN component. The GSCC is the intersection of the IN and OUT components. All sites in the GWCC, but not in the IN and OUT components, are referred to as the "tendrils" (see Fig. 1).

For a directed random network of arbitrary degree distribution the condition for the existence of a giant component can be deduced in a manner similar to [9]. If a site, *b*, is reached following a link pointing to it from site *a*, then it must have at least one outgoing link, on average, in order to be part of a giant component. This condition can be written as

$$\langle k_b | a \to b \rangle = \sum_{j_b, k_b} k_b P(j_b, k_b | a \to b) = 1, \tag{2}$$

where *j* and *k* are the in and out degrees, respectively, $P(j_b, k_b | a \to b)$ is the conditional probability given that site *a* has a link leading to *b*, and $\langle k_b | a \to b \rangle$ is the conditional average. Using Bayes rule we get

$$P(j_b, k_b | a \to b) = P(j_b, k_b, a \to b)/P(a \to b)$$
$$= P(a \to b | j_b, k_b) P(j_b, k_b)/P(a \to b).$$

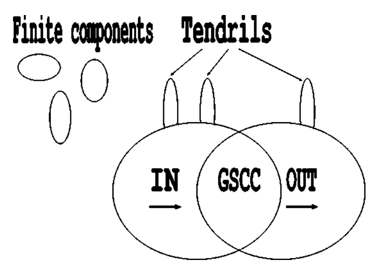

FIG. 1. Structure of a general directed graph.

For random networks $P(a\to b)=\langle k\rangle/(N-1)$ and $P(a\to b|j_b,k_b)=j_b/(N-1)$, where N is the total number of nodes in the network. The above criterion thus reduces to [3,4]

$$\langle jk\rangle \geqslant \langle k\rangle. \tag{3}$$

Suppose a fraction p of the nodes is removed from the network. (Alternatively, a fraction $q=1-p$ of the nodes is retained.) The original degree distribution, $P(j,k)$, becomes

$$P'(j,k)=\sum_{j_0,k_0}^{\infty} P(j_0,k_0)\binom{j_0}{j}(1-p)^j p^{j_0-j} \times\binom{k_0}{k}(1-p)^k p^{k_0-k}. \tag{4}$$

In view of this new distribution, Eq. (3) yields the percolation threshold

$$q_c=1-p_c=\frac{\langle k\rangle}{\langle jk\rangle}, \tag{5}$$

where averages are computed with respect to the original distribution before dilution, $P(j,k)$. Equation (5) indicates that in directed scale-free networks if $\langle jk\rangle$ diverges then $q_c\to 0$ and the network is resilient to random breakdown of nodes and bonds.

The term $\langle jk\rangle$ may be dramatically influenced by the appearance of correlations between the in and out degrees of the nodes. In particular, let us consider scale-free distributions for both the in and out degrees

$$P_{in}(j)\sim\begin{cases} Bc_{in}j^{-\lambda_{in}}, & j\neq 0\\ 1-B, & j=0\end{cases} \tag{6}$$

and

$$P_{out}(k)=c_{out}k^{-\lambda_{out}}. \tag{7}$$

In Eq. (6) we choose to add the possible zero value to the in degree in order to maintain $\langle j\rangle=\langle k\rangle$. If the in and out degrees are uncorrelated, we expect $\langle jk\rangle=\langle j\rangle\langle k\rangle$. For several real directed networks this equality does not hold. For example, the network of Notre Dame University WWW [6] has $\langle k\rangle=\langle j\rangle\approx 4.6$, and thus $\langle j\rangle\langle k\rangle=21.16$. In contrast, measuring directly we find $\langle jk\rangle\approx 200$, about an order of magnitude larger than the result expected for the uncorrelated case. This yields an estimate of $q_c\approx 0.02$, i.e., a very stable directed network. We obtained similar results also for some metabolic networks [16], indicating that in real directed networks, the in and out degrees are correlated.

To address correlations, we model it in the following manner: we first generate the j values for the entire network. Next, for each site with $j\neq 0$ with probability A we generate k fully correlated with j, i.e., $k=k(j)$. Assuming that $k(j)$ is a monotonically increasing function then the requirement $c_{out}k^{-\lambda_{out}}dk=c_{in}j^{-\lambda_{in}}dj$—needed to maintain the distributions scale-free—leads to $k^{\lambda_{out}-1}=j^{\lambda_{in}-1}$. With probability $1-A$, the degree k is chosen independently from j,

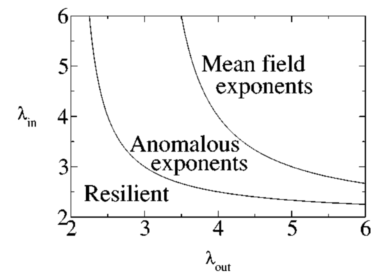

FIG. 2. Phase diagram of the different regimes for the IN component of scale-free correlated directed networks. The boundary between resilient and anomalous exponents is derived from Eq. (9) while that between anomalous exponents and mean field exponents is given by Eq. (24) for $\lambda^\star=4$. For the diagram of the OUT component λ_{in} and λ_{out} change roles.

$$P(j,k)\sim\begin{cases}(1-A)Bc_{in}j^{-\lambda_{in}}c_{out}k^{-\lambda_{out}}\\ +BAc_{out}k^{-\lambda_{out}}\delta_{j,j(k)}, & j\neq 0\\ (1-B)c_{out}k^{-\lambda_{out}}, & j=0,\end{cases} \tag{8}$$

where $j(k)=k^{(\lambda_{out}-1)/(\lambda_{in}-1)}$. With this distribution, any finite fraction BA of fully correlated sites yields a diverging $\langle jk\rangle$ whenever

$$(\lambda_{out}-2)(\lambda_{in}-2)\leqslant 1, \tag{9}$$

causing the percolation threshold to vanish (see Fig. 2).

In the case of no correlations between the in and the out degrees, $A=0$, Eq. (8) becomes $P(j,k)=P_{in}(j)P_{out}(k)$. Then the condition for the existence of a giant component is $\langle k\rangle=\langle j\rangle=1$. Moreover, Eq. (5) reduces to

$$q_c=1-p_c=\frac{1}{\langle k\rangle}. \tag{10}$$

Applying Eq. (10) to scale-free networks one concludes that for $\lambda_{out}>2$ and $\lambda_{in}>2$ a phase transition exists at a finite q_c. Here we concern ourselves with the critical exponents associated with the percolation transition in scale-free network of $\lambda_{out}>2$ and $\lambda_{in}>2$ which is the most relevant regime (Fig. 2).

Percolation of the GWCC can be seen to be similar to percolation in the non-directed graph created from the directed graph by ignoring the directionality of the links. The threshold is obtained from the criterion [9]

$$q_c=\frac{\langle k\rangle}{\langle k(k-1)\rangle}. \tag{11}$$

Here the connectivity distribution is the convolution of the in and out distributions

$$P'(k)=\sum_{l=0}^{k} P(l,k-l). \tag{12}$$

Regardless of correlations, $P'(k)$ is always dominated by the slower decay exponent, therefore percolation of the GWCC is the same as in nondirected scale-free networks, with $\lambda_{eff}=\min(\lambda_{in},\lambda_{out})$. Note that the percolation threshold of the GWCC may differ from that of the GSCC and the IN and OUT components [4].

We now use the formalism of generating functions [17,18] to analyze percolation of the GSCC and IN and OUT components. In [3,4] a generating function is built for the joint probability distribution of outgoing and incoming degrees, before dilution:

$$\Phi(x,y)=\sum_{k,j} P(j,k)x^j y^k. \tag{13}$$

Using the approach of Callaway *et al.* [10] let $q(j,k)$ be the probability that a vertex of degree (j,k) remains in the network following dilution. The generating function after dilution is then

$$G(x,y)=\sum_{k,j} P(j,k)q(j,k)x^j y^k. \tag{14}$$

From Eq. (14) it is possible to define the generating function for the outgoing degrees G_0:

$$G_0(y)\equiv G(1,y)=\sum_{k,j} P(j,k)q(j,k)y^k. \tag{15}$$

The probability of reaching a site by following a specific link is proportional to $jP(j,k)$, therefore, the probability of reaching an occupied site following a specific directed link is generated by

$$G_1(y)=\frac{\sum_{j,k} jP(j,k)q(j,k)y^k}{\sum_{j,k} jP(j,k)}. \tag{16}$$

Let $H_1(y)$ be the generating function for the probability of reaching an outgoing component of a given size by following a directed link, after a dilution. $H_1(y)$ satisfies the self-consistent equation:

$$H_1(y)=1-G_1(1)+yG_1(H_1(y)). \tag{17}$$

Since $G_0(y)$ is the generating function for the outgoing degree of a site, the generating function for the probability that n sites are reachable from a given site is

$$H_0(y)=1-G_0(1)+yG_0(H_1(y)). \tag{18}$$

For the case where correlations exist, and assuming random dilution: $q(j,k)=q$, Eqs. (17) and (18) reduce to

$$H_1(y)=1-q+\frac{qy}{\langle j\rangle}\sum_k [BAj(k)+(1-A)\times\langle j\rangle]P_{out}(k)H_1(y)^k \tag{19}$$

and

$$H_0(y)=1-q+qy\sum_k P_{out}(k)H_1(y)^k. \tag{20}$$

If $A\to 0$, one expects that $H_0(y)=H_1(y)$, since there is no correlation between j and k, thus the probability to have k outgoing edges is $P_{out}(k)$ whether we choose the site randomly or weighted by the incoming edges j.

$H_0(1)$ is the probability to reach an outgoing component of any *finite* size choosing a site. Thus, below the percolation transition $H_0(1)=1$, while above the transition there is a finite probability to follow a directed link to a site which is a root of an infinite outgoing component: $P_\infty=1-H_0(1)$. It follows that

$$P_\infty(q)=q\left(1-\sum_k^\infty P_{out}(k)u^k\right), \tag{21}$$

where $u\equiv H_1(1)$ is the smallest positive root of

$$u=1-q+\frac{q}{\langle j\rangle}\sum_k [BAj(k)+(1-A)\langle j\rangle]P_{out}(k)u^k. \tag{22}$$

Here $P_\infty(q)$ is the fraction of sites from which an infinite number of sites is reachable. Equation (22) can be solved numerically and the solution may be substituted into Eq. (21), yielding the size of the IN component at dilution $p=1-q$.

Near criticality, the probability to start from a site and reach a giant outgoing component follows $P_\infty\sim(q-q_c)^\beta$. For mean-field systems (such as infinite-dimensional systems, random graphs, and Cayley trees) it is known that $\beta=1$ [19]. This regular mean-field result is not always valid. Instead, following [20] we study the behavior of Eq. (22) near $q=q_c, u=1$, and find

$$\beta=\begin{cases}\dfrac{1}{3-\lambda^\star}, & 2<\lambda^\star<3\\[2ex] \dfrac{1}{\lambda^\star-3}, & 3<\lambda^\star<4\\[2ex] 1, \quad \lambda^\star>4, & \end{cases} \tag{23}$$

where

$$\lambda^\star=\lambda_{out}+\frac{\lambda_{in}-\lambda_{out}}{\lambda_{in}-1}. \tag{24}$$

We see that the order parameter exponent β attains its usual mean-field value only for $\lambda^\star>4$. As $\lambda_{out}\to\lambda_{in}$ the correlated fraction BA of sites resembles nondirected networks [20,21] (where there is no distinction between incoming and outgoing degrees). In this case we get $\lambda^\star=\lambda_{out}=\lambda_{in}$ for any amount of correlation A. The criterion for the existence of a giant component is then $\langle k^2\rangle/\langle k\rangle=1$, and not 2 as in the nondirected case. The difference stems from the fact that in the nondirected case one of the links is used to reach the site, while in the directed case there is generally no correlation between the location of the incoming and outgoing links. Therefore, one more outgoing link is available for leaving the site.

Without any correlations, $A=0$, different terms prevail in the analysis and

$$\beta=\begin{cases}\dfrac{1}{\lambda_{out}-2}, & 2<\lambda_{out}<3\\[2ex] 1, \quad \lambda_{out}>3. & \end{cases} \tag{25}$$

TABLE I. Values of $\lambda^\star$ for the different network components for both correlated and uncorrelated cases.

	Uncorrelated	Correlated
GWCC	$\min(\lambda_{out},\lambda_{in})+1$	$\min(\lambda_{out},\lambda_{in})$
IN	$\lambda_{out}+1$	$\lambda_{out}+\frac{\lambda_{in}-\lambda_{out}}{\lambda_{in}-1}$
OUT	$\lambda_{in}+1$	$\lambda_{in}+\frac{\lambda_{out}-\lambda_{in}}{\lambda_{out}-1}$
GSCC	$\min(\lambda_{out},\lambda_{in})+1$	$\min(\lambda_{out}^{*},\lambda_{in}^{*})$

This is the same as Eq. (23) but with $\lambda^\star=\lambda_{out}+1$.

The GSCC is the intersection of the IN and OUT components. Therefore, it behaves as the smaller of the two components: $\beta_{GSCC}=\max(\beta_{in},\beta_{out})$. This can be also derived by applying the same methods as for the IN and OUT components to the generating function of the GSCC obtained in [4]. The exponent for the GWCC, on the other hand, is independent of the exponents of the other components, since the transition point is different.

It is known that for a random graph of arbitrary degree distribution the finite clusters follow the scaling form

$$n(s)\sim s^{-\tau}e^{-s/s^*}, \tag{26}$$

where s is the cluster size and $n(s)$ is the number of clusters of size s. At criticality $s^*\sim|q-q_c|^{-\sigma}$ diverges and the tail of the distribution follows a power law.

The probability that s sites can be reached from a site by following links at criticality follows $p(s)\sim s^{-\tau}$, and is generated by H_0, where $H_0(y)=\Sigma_s p(s)y^s$. As in [20], $H_0(y)$ can be expanded from Eq. (18). In the presence of correlations we find

$$\tau=\begin{cases} 1+\dfrac{1}{\lambda^\star-2}, & 2<\lambda^\star<4 \\ \dfrac{3}{2}, & \lambda^\star>4. \end{cases} \tag{27}$$

The regular mean-field exponents are recovered for $\lambda^\star>4$. For the uncorrelated case we get

$$\tau=\begin{cases} 1+\dfrac{1}{\lambda_{out}-1}, & 2<\lambda_{out}<3 \\ \dfrac{3}{2}, & \lambda_{out}>3. \end{cases} \tag{28}$$

Now the regular mean-field results are obtained for $\lambda>3$.

In summary, we calculate the percolation properties of directed scale-free networks. We find that the percolation critical exponents in scale-free networks are strongly dependent upon the existence of correlations and upon the degree distribution exponents in the range of $2<\lambda^\star<4$. This regime characterizes most naturally occurring networks, such as metabolic networks or the WWW. The regular mean-field behavior of percolation in infinite dimensions is recovered only for $\lambda^\star>4$. A connection is found between nondirected and directed scale-free percolation exponents for any finite correlation between the in and out degrees. In the uncorrelated case, i.e. $P(j,k)=P_{in}(j)P_{out}(k)$, the probability to reach an outgoing component does not bear any dependence upon $P_{in}(j)$. The results are summarized in Table I.

Support from the NSF is gratefully acknowledged (D.bA).

[1] R. Albert and A.L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. E **64**, 026118 (2001).
[4] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. E **64**, 025101R (2001).
[5] A.D. Sánchez, J.M. López, and M.A. Rodríguez, Phys. Rev. Lett. **88**, 048701 (2002).
[6] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **281**, 2115 (2000).
[7] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).
[8] R. Albert, H. Jeong, and A.L. Barabási, Nature (London) **406**, 6794 378 (2000); **406**, 378 (2000).
[9] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[10] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[11] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[12] R.V. Sole and J.M. Montoya, Proc. R. Soc. London, Ser. B **268**, 2039 (2001).
[13] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor and Francis, London, 1991).
[14] *Fractals and Disordered System*, edited by A. Bunde, and S. Havlin (Springer, New York, 1996).
[15] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).
[16] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási, Nature (London) **407**, 651 (2000).
[17] H.S. Wilf, *Generatingfunctionology*, 2nd ed. (Academic Press, London, 1994).
[18] G.H. Weiss, *Aspects and Applications of the Random Walk* (North-Holland, Amsterdam, 1994).
[19] P. Frojdh, M. Howard, and K.B. Lauritsen, Int. J. Mod. Phys. B **15**, 1761 (2001).
[20] R. Cohen, D. ben-Avraham, and S. Havlin, e-print cond-mat/0202259.
[21] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).