Percolation and blind spots in complex networks

Liang Huang, Ying-Cheng Lai, Kwangho Park, and Junshan Zhang
Department of Electrical Engineering, Arizona State University, Tempe, Arizona 85287, USA

(Received 7 February 2006; published 30 June 2006)

Recent works on network security have focused on whether a complex network can maintain its integrability under attack or random node failures. In applications of increasing importance such as sensor networks, a somewhat different problem, namely, the occurrence of isolated nodes (or blind spots), is of great interest. We show that, for networks with a stronger ability to form global spanning clusters, it is relatively more difficult to eliminate blind spots, and vice versa. We use the framework of percolation to investigate this phenomenon. Our analysis also yields a formula for the average number of blind spots, which provide an explanation for several numerical findings.

DOI: 10.1103/PhysRevE.73.066131 PACS number(s): 89.75.Hc, 89.20.Hh, 89.75.Da

The ability for a complex network to maintain its integrability in response to attacks or random failures has been an issue of tremendous interest [1–12] since the discoveries of the small-world [13] and the scale-free [14] topologies. Loosely, a network can be considered as integrable and functional if a substantial fraction of nodes are connected. Theoretically, the problem can be treated in the framework of percolation where one can ask, for instance, under what conditions a global spanning cluster of nodes—which contains a considerable fraction of the active nodes—can be formed [2,3,15]. In particular, consider a network embedded in a lattice with a prescribed degree distribution. The extent of an attack or random failures can be conveniently measured by the occupying probability of active nodes where, for instance, a severe attack that disables a large number of nodes corresponds to a small occupying probability of active nodes on the lattice. One can then ask whether a threshold value of the probability exists, above which a spanning cluster is formed. The pioneering works by Cohen *et al.* [2] and by Callaway *et al.* [3] indicated that for networks with random linkage, if the second moment of the degree distribution diverges, the percolation threshold tends to zero. It has been found subsequently that other network properties such as the degree correlation [16,17], degree of clustering [18], and the geometry of lattice embedding [19,20] etc., can also affect the percolation threshold. For instance, they can lead to a nonzero threshold even when the second moment in the degree distribution diverges.

There are applications of increasing importance and wide interest, such as sensor networks or multihop ad hoc networks, in which a central issue of concern is the occurrence of isolated nodes, or blind spots. For instance, for a cellular network, total disintegration of the network is rare, i.e., whether there is a spanning cluster is not an issue. What one concerns most is perhaps whether individual customers can get access to the network. In a sensor network for homeland defense applications, the occurrence of blind spots indicates the loss of information that may be critical for certain tasks. Intuitively, one may expect that networks with a stronger ability to form spanning clusters should be more capable of "absorbing" isolated nodes and, hence, such networks should be more robust against the occurrence of blind spots. In the language of percolation, this is to say that networks with smaller percolation thresholds should be more fully connected as the occupying probability is increased beyond the threshold. What we find in this paper is the opposite: *blind spots are more probable in networks that are more susceptible to percolation*. This striking phenomenon may present a significant challenge to the design of secure and reliable networks; to make the network robust against attacks or random failures, it is necessary to reduce the percolation threshold, but the network may be unreliable from the standpoint of users because of the relatively higher likelihood of blind spots. In the following we shall present a systematic numerical study of the behavior of the blind spots and provide a theoretical analysis.

We begin by reviewing the continuous process of percolation where N nodes of a network fill in sites on a "lattice," where each node is occupied (or active) with probability q. Thus $p \equiv 1-q$ is the probability that a site is empty or the probability of node failure. On average there are Nq nodes on the network. Assume that the final network is scale-free, i.e., the degrees of the nodes follow the power-law distribution $k^{-\lambda}$. As q is increased, more and more sites are occupied so that nodes begin to connect with each other to form local clusters of various sizes. When q is increased beyond the percolation threshold q_t, a global spanning cluster emerges which connects nodes across the entire lattice [15]. For networks with random linkage in the connection topology, the threshold is given by $\langle k\rangle/(\langle k^2\rangle-\langle k\rangle)$, where $\langle k\rangle$ and $\langle k^2\rangle$ are the average degree and the second moment of the degree distribution, respectively [2,3]. Assume that the degree-distribution exponent satisfies $\lambda>3$ so that $\langle k\rangle$ and $\langle k^2\rangle$ exist. In this case, the more heterogeneous the network, the larger the second moment $\langle k^2\rangle$, and the smaller the threshold. For $2<\lambda\leqslant 3$, $\langle k^2\rangle$ diverges while $\langle k\rangle$ exists, thus the threshold is 0 in the large network limit. However, if the linkage is not random but has structures, the percolation threshold can still be nonzero [16–20]. In these cases, as Warren *et al.* have shown [19], it is also true that the more heterogeneous the network, the smaller the threshold. The general result is that percolation occurs more easily for networks with smaller values of λ.

Generally, as q is increased, the number of blind spots decreases. To obtain an understanding of the behavior of blind spots, it is convenient to use statistical quantities. Let $\langle n_s\rangle$ be the ensemble-averaged value of the number of

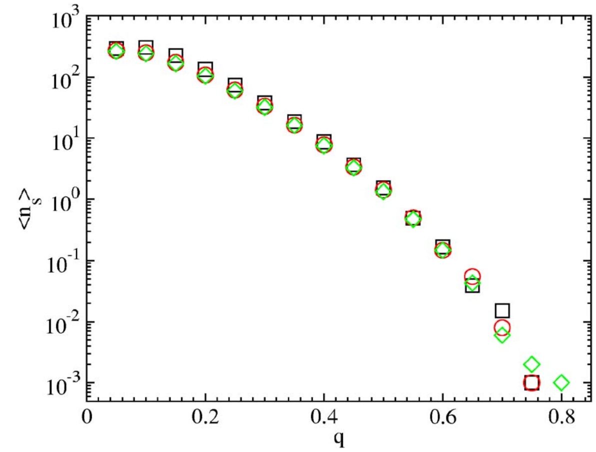

FIG. 1. (Color online) The average number of blind spots $\langle n_s \rangle$ versus the occupying probability q for three scale-free networks of $N=10^4$ nodes with the same degree distribution k^{-3} but generated using the lattice-embedding model (squares), the random linkage model (circles), and the classical preferential-attachment model (diamonds). The minimal value of the degree is $m=10$. Each data point is the result of averaging over 10^3 network realizations. While these models can generate different higher order, fine statistical properties for the network, the behavior of the blind spots is apparently independent of the fine details.

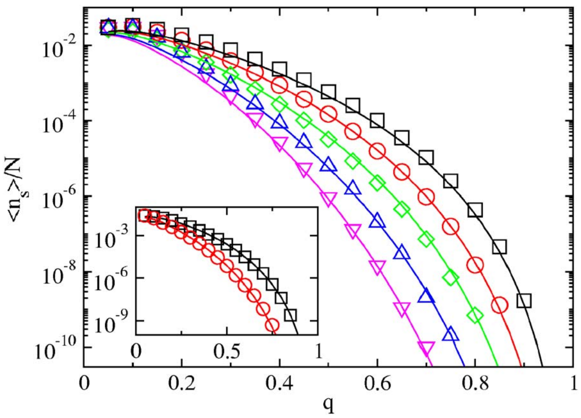

FIG. 2. (Color online) Normalized average number of blind spots $\langle n_s \rangle / N$ versus the occupying probability q for four scale-free networks of $N=10^4$ with approximately the same average degree: $\lambda=2.6$, $m=8$ (squares), $\lambda=3$, $m=10$ (circles), $\lambda=3.5$, $m=12$ (diamonds), $\lambda=5$, $m=15$ (up triangles), and $\lambda=30$, $m=19$ (down triangles). Each data point is the result of averaging over 10^6 network realizations. Solid curves are from theoretical prediction Eq. (1) (not fits). Inset: the same quantities and statistics for optimal two-peak networks with $\langle k \rangle=20$ for $k_l=10$, $k_m=30$ (squares), and $k_l=15$, $k_m=25$ (circles).

isolated finite clusters (or blind spots), which can assume noninteger values. Whether the network is fully connected, i.e., no blind spots, can then be determined by the criterion that $\langle n_s \rangle$ falls below one. This defines another threshold value q_c, the critical value of the occupying probability required for the network to be fully connected. Figure 1 shows $\langle n_s \rangle$ versus q for three scale-free networks obtained by different models: the lattice-embedding model [21], random scale-free model [22], and the classical preferential-attachment model [14]. An observation is that the behavior of $\langle n_s \rangle$ appears to depend on the degree distribution only, regardless of the model details that can give rise to distinct fine, higher-order statistical properties for the network. This should be contrasted with the percolating process to form a spanning cluster, which is sensitive to fine network properties such as the degree correlation [16,17], clustering [18], lattice embedding methods [19,20], etc.

The particular physical networks for which the problem of blind spots may be of concern are sensor networks, multihop wireless networks, and possibly the internet. To build up numerical models for these networks, taking into account the effect of physical distances, the lattice-embedding method [21] is appropriate. By this method, for each node on an $L \times L$ lattice with periodic boundaries, we assign a degree k, drawn from the scale-free distribution of exponent λ and minimal degree m. We then randomly choose a node and connect it to its neighbors from near to far, within the distance $A\sqrt{k}$ (e.g., this distance can be adjusted by power control in wireless networks), where A is a prescribed constant and the distance between two nearest-neighbor nodes in the lattice is defined to be one. The process is repeated for all nodes. If A is suitably large (e.g., $A=7$ for $N=10^4$), the degree distribution can be realized. In the following we shall use different scale-free networks generated by this model to explore the behavior of the blind spots and to contrast it with percolation.

Figure 2 shows the dependence of the normalized number of blind spots $\langle n_s \rangle / N$ on the occupying probability q for scale-free networks with different parameters, which are chosen so that all the networks have approximately the same value of the average degree: $\langle k \rangle \approx 20$, for a meaningful comparison (e.g., the cost to build up all the network is the same, if it is proportional to the total number of links). We see that for networks with larger values of λ (or relatively more homogeneous networks), the number of blind spots decreases more rapidly with q, indicating that, under the same cost, scale-free networks with larger values of the degree exponent are structurally more robust in the sense of minimizing the number of blind spots. The solid curves are not fits but are predictions of our theoretical formula Eq. (1) (to be described below).

To compare the behavior of the blind spots with the percolation process, we focus on the two probabilities as a function of the occupying probability: $\Phi_{sp}(q)$, the probability to form a spanning cluster through percolation, defined as N'/Nq, where N' is the size of the largest cluster; and $\Phi_{fc}(q)$, the probability for the network to be fully connected. Some representative results are shown in Figs. 3(a) and 3(b) where the open symbols are for $\Phi_{sp}(q)$ and the filled symbols are for $\Phi_{fc}(q)$, and all networks have the same number of nodes. In Fig. 3(a), the average degree is approximately 20 for all five networks. It can be seen that, although the formation of a spanning cluster requires slightly smaller value of q for relatively more heterogeneous networks (e.g., the curve with $\lambda=2.6$ represented by open squares), the values of q required for the network to be fully connected are much larger. That is, it is more difficult to eliminate blind spots for

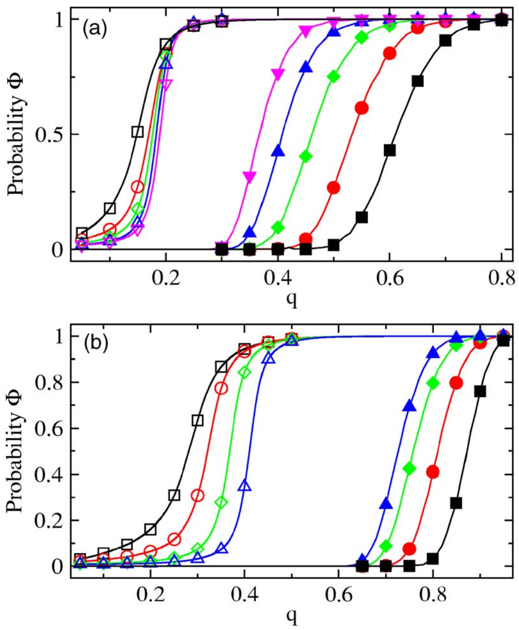

FIG. 3. (Color online) Probability $\Phi_{sp}(q)$ (open symbols) for a spanning cluster to be formed and the probability $\Phi_{fc}(q)$ (filled symbols) for the network to be fully connected, for two sets of scale-free networks, each of $N=10^4$ nodes, (a) $\langle k\rangle \approx 20$: $\lambda=2.6$, $m=8$ (squares), $\lambda=3$, $m=10$ (circles), $\lambda=3.5$, $m=12$ (diamonds), $\lambda=5$, $m=15$ (up triangles), and $\lambda=30$, $m=19$ (down triangles); (b) $\langle k\rangle \approx 10$: $\lambda=2.6$, $m=4$ (squares), $\lambda=3$, $m=5$ (circles), $\lambda=4$, $m=6$ (diamonds), and $\lambda=9$, $m=7$ (triangles). The values of $\Phi_{sp}(q)$ and $\Phi_{fc}(q)$ are calculated using 10^3 and 10^4 network realizations, respectively.

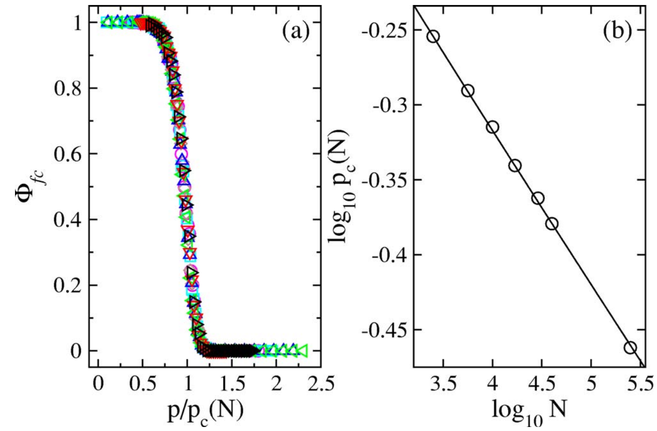

FIG. 4. (Color online) For a scale-free network of $\lambda=3$ and $m=10$, (a) universal behavior in the fully-connected probability Φ_{fc} versus the normalized occupying probability $p/p_c(N)$ for seven different network sizes: 2 500, 5 625, 10 000, 16 900, 28 900, 40 000, and 250 000, where each data point is the result of an ensemble average of 1000 networks, and (b) $\log p_c(N)$ versus $\log N$. The solid line in (b) is calculated from the theoretical formula Eq. (1) (not a fit).

networks that are more susceptible to percolation! This contrast is more striking for networks with smaller average degrees, as shown in Fig. 3(b), where $\langle k\rangle \approx 10$. Qualitatively, this can be understood by noting that, although the percolation threshold is generally smaller for relatively more heterogeneous networks, when the average degree is fixed, there is also a higher probability for these networks to possess more small-degree nodes, making a full connection more difficult.

For an infinite network, the critical occupying probability for the disappearance of blind spots is $q_c=1$ (or $p_c=0$). Of physical importance are finite-size effects. We are thus led to ask the following question: How does q_c (or p_c) scale with the system size N? To address this question, it is convenient to focus on the dependence $\Phi_{fc}(p)$. For $p\to 0$ (or $q\to 1$) we have $\Phi_{fc}(p)=1$ and it becomes zero for $p\to 1$. Thus a transition in $\Phi_{fc}(p)$ from one to zero occurs around p_c. This transition point naturally depends on the system size (e.g., for $N\to\infty$, $p_c\to 0$), so we write $p_c(N)$. For networks with the same degree distribution but with different sizes, it is reasonable to assume that the curves $\Phi_{fc}(p)$ are translated versions of one another, where the curves for larger systems are shifted more toward the left (i.e., $p=0$). These arguments suggest that, if we use the rescaled variable $p/p_c(N)$, then all probabilities $\Phi_{fc}[p/p_c(N)]$ should collapse into a single, universal curve, irrespective of the system size. This has indeed been observed numerically, as shown in Fig. 4(a). From the various system sizes used in Fig. 4(a), we obtain the dependence $p_c(N)$, as shown by the open circles in Fig. 4(b). We observe that $p_c(N)\sim N^{-\alpha}$. The solid line in Fig. 4 is not a fit but is directly calculated by setting the number of blind spots $n_s=1$ in our theoretical formula Eq. (1).

We now provide a heuristic approach to explain the numerical results. To analyze the occurrence of the blind spots, we note that, a node of degree k is isolated if it is occupied but all k sites connecting to it are unoccupied. The probability of this event is qp^k. Let $P(k)$ be the degree distribution. Then, on average, the total number of single-node blind spots is $n_1=N\Sigma_k qp^k P(k)$. Similarly, the average number of 2-node blind spots is $n_2=\Sigma q^2 p^{k_1+k_2-2} N_e P(k_1,k_2)$, where the summation is over all sorted pairs of k_1 and k_2, $N_e=N\langle k\rangle/2$ is the number of edges, and $P(k_1,k_2)$ is the joint degree distribution. The number of m-node blind spots can be obtained similarly by summing over different configurations, which is proportional to $p^{k_1+k_2+\cdots+k_m}$. Since we are only interested in the case where q is far beyond q_t (the percolation threshold), the numbers of higher-order blind spots are negligible (Fig. 2) as compared to the number of single-node blind spots. For a scale-free network of minimal degree m, the degree distribution can be written as $P(K)=Ck^{-\lambda}$, for $k\geq m$, where C is a normalization constant given by $C=1/\Sigma_{k=m}^{\infty}k^{-\lambda}=1/\zeta(\lambda,m)$, and $\zeta(\lambda,m)$ is the Hurwitz Zeta function [23]. We thus have, for the number of blind spots,

$$n_s=\sum_{k=m}^{\infty} Nqp^k Ck^{-\lambda} = Nqp^m C\sum_{k=0}^{\infty}\frac{p^k}{(k+m)^{\lambda}} = Nqp^m Li_{\lambda}(p,m)/\zeta(\lambda,m), \tag{1}$$

where $Li_{\lambda}(p,m)$ is the generalized polylogarithm function, and $Li_{\lambda}(1,m)=\zeta(\lambda,m)$ [23]. Equation (1) is valid for $\lambda>1$. As $q\to 1$ (or $p\to 0$), we have $n_s/N\to 0$. Thus for a fixed network size N, there exists a critical value q_c of the probability q such that for $q > q_c$ (or $p < p_c$), the number of blind spots falls below one so that the network is fully connected. The smaller the q_c value, the more easily the network can be fully connected. The critical value can be obtained by setting $n_s = 1$ [e.g., Fig. 4(b)].

To explain the algebraic relation between $p_c(N)$ and N, we consider the large λ cases where $Li_\lambda(p,m)/\zeta(\lambda,m)$ tends to one, thus the number of blind spots becomes $n_s \sim Nqp^m \lesssim Np^m$. We then have

$$p_c \sim N^{-\alpha}, \tag{2}$$

where $\alpha = 1/m$ is the algebraic scaling exponent. In general, there is no closed form of $Li_\lambda(p,m)$, so a precise scaling law cannot be obtained explicitly. However, since $Li_\lambda(p,m) = \frac{1}{m^\lambda} + \frac{p}{(1+m)^\lambda} + \cdots \simeq m^{-\lambda}\left(1 + \left(\frac{m}{1+m}\right)^\lambda p\right)$, for small p, we can expect that the scaling of p_c is similar to Eq. (2), but has a small correction, due to the term of $\left(1 + \left(\frac{m}{1+m}\right)^\lambda p\right)$. This correction diminishes for large λ. A linear fit of the data in Fig. 4(b) gives the numerical value of the algebraic scaling exponent as $\alpha \approx 0.1$, which agrees with the theoretical value of $1/m = 0.1$ quite well.

For random bond failures, let p_b be the failure probability of a bond and $q_b = 1 - p_b$ be the occupying probability. A node of degree k will be isolated if all the k bonds connecting to it are failed, and the probability of this event is p_b^k. Again, since the critical point of the occurrence of blind spots q_{bc} is far beyond the percolation threshold q_{bt}, we can neglect the higher order blind spots. The number of blind spots is then $n_s = Np_b^m Li_\lambda(p_b,m)/\zeta(\lambda,m)$. Letting $n_s = 1$, we can obtain the critical point p_{bc} and the scaling relation $p_{bc} \sim N^{-1/m}$. This analysis has been verified numerically.

To further validate the generality of our result, we have studied the class of two-peak networks defined by the following degree distribution: $P(k) = 0$ if $k \neq k_l$ or $k \neq k_m$ for $k_l \leqslant k_m$, where k_l and k_m are the minimal and maximal degree of the network, respectively. The values of k_l and k_m are chosen to optimize the network's robustness to random node failures for a given average degree $\langle k \rangle$ [11,12]. A similar calculation yields $n_s = Nq\{p^{k_l}(k_m - \langle k \rangle)/(k_m - k_l) + p^{k_m}(\langle k \rangle - k_l)/(k_m - k_l)\}$. The inset of Fig. 2 shows the simulation result (symbols) and the theory (curves) of two cases with $\langle k \rangle = 20$: $k_l = 10$, $k_m = 30$ (squares) and $k_l = 15$, $k_m = 25$ (circles). Since the range $[k_l, k_m]$ for the second case is contained entirely in that of the first case, the analysis of Refs. [11,12] implies a higher percolation threshold for the second case. At the same time, the second case has fewer blind spots, as our computation reveals. The conclusion is that our finding holds for the optimal two-peak networks as well.

In summary, we have investigated the behavior of the blind spots in relation to the percolation process in scale-free networks. Numerical findings and analysis indicate that a heterogeneous degree distribution, while important for making the network function by facilitating the emergence of a spanning cluster through a small subset of high-degree nodes, may at the same time cause difficulty for small-degree nodes to be connected. This "full connectivity" may be particularly important for a variety of situations [24–26] such as sensor networks, multihop ad hoc networks, small-scale intercomputer networks in business or defense applications, or even the internet [27]. There has not been much attention to this problem, and we hope our work will stimulate more efforts in this direction.

This work is supported by NSF under Grant No. ITR-0312131 and by AFOSR under Grants No. F49620-01-01-0317 and No. FA9550-06-1-0024, and by ONR under Grant No. N00014-05-1-0636.

[1] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[2] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000); **86**, 3682 (2001).
[3] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[4] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[5] P. Holme and B. J. Kim, Phys. Rev. E **65**, 066109 (2002).
[6] Y. Mereno, J. B. Gómez, and A. F. Pacheco, Europhys. Lett. **58**, 630 (2002); Y. Moreno, R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *ibid.* **62**, 292 (2003).
[7] P. Holme, Phys. Rev. E **66**, 036119 (2002).
[8] A. E. Motter and Y.-C. Lai, Phys. Rev. E **66**, 065102(R) (2002); L. Zhao, K. Park, and Y.-C. Lai, *ibid.* **70**, 035101(R) (2004); L. Zhao, K. Park, Y.-C. Lai, and N. Ye, *ibid.* **72**, 025104(R) (2005).
[9] K.-I. Goh, C.-M. Ghim, B. Kahng, and D. Kim, Phys. Rev. Lett. **91**, 189804 (2003).
[10] A. E. Motter, Phys. Rev. Lett. **93**, 098701 (2004).
[11] A. X. C. N. Valente, A. Sarkar, and H. A. Stone, Phys. Rev. Lett. **92**, 118702 (2004).
[12] G. Paul, S. Sreenivasan, and H. E. Stanley, Phys. Rev. E **72**, 056130 (2005); G. Paul, S. Sreenivasan, S. Havlin, and H. E. Stanley, Physica A (to be published).
[13] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[14] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[15] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).
[16] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[17] A. Vazquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).
[18] E. Volz, Phys. Rev. E **70**, 056115 (2004).
[19] C. P. Warren, L. M. Sander, and I. M. Sokolov, Phys. Rev. E **66**, 056105 (2002).
[20] L. Huang, L. Yang, and K. Yang, Europhys. Lett. **72**, 144 (2005); e-print physics/0503147 (2005).
[21] A. F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **89**, 218701 (2002).
[22] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[23] H. M. Srivastava and J. Choi, *Series Associated with the Zeta*

and Related Functions (Kluwer Academic, Dordrecht, 2001).
[24] M. D. Penrose, Ann. Appl. Probab. **7**, 340 (1997).
[25] P. Gupta and P. R. Kumar, in *Stochastic Analysis, Control, Optimization and Applications*, edited by W. M. McEneaney, G. Yin, and Q. Zhang (Birkhäuser, Boston, 1998).
[26] M. Franceshetti and R. Meester, "Critical node lifetimes in random networks via the Chen-Stein method," in *Proceedings of ISIT 2005*.
[27] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).