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 multipath 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 = 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−λ. 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 qc, 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 (k)/((k2)−(k)), where (k) and (k2) are the average degree and the second moment of the degree distribution, respectively [2,3]. Assume that the degree-distribution exponent satisfies λ > 3 so that (k) and (k2) exist. In this case, the more heterogeneous the network, the larger the second moment (k2), and the smaller the threshold. For 2 < λ ≤ 3, (k2) diverges while (k) 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 homogeneous 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 ⟨n⟩ be the ensemble-averaged value of the number of
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 \( \lambda \) 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 \( N=10^5 \)), 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: \( \langle k \rangle = 20 \), 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 \( \lambda \) (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 \( N'/N_q \), 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
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 \( \langle k \rangle = 10 \). 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 occupancy 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 \( p_c(N) \to N^{-\alpha} \). 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 \( q^k \). Let \( P(k) \) be the degree distribution. Then, on average, the total number of single-node blind spots is \( n_s = N \sum q^k P(k) \). Similarly, the average number of 2-node blind spots is \( n_s^2 = N^2 \sum q^k P(k) \). Since we are only interested in the case where \( q \) is far beyond \( q_c \) (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 \( P(k) = Ck^{-\alpha} \), for \( k \geq m \), where \( C \) is a normalization constant given by \( C = 1/\sum_{k=m}^{\infty} k^{-\alpha} = 1/\zeta(\lambda, m) \), and \( \zeta(\lambda, m) \) is the Hurwitz Zeta function [23]. We thus have, for the number of blind spots,

\[
 n_s = \sum_{k=m}^{\infty} Nq^k Ck^{-\alpha} = Nq^m C \sum_{k=0}^{\infty} \frac{p^k}{(k + m)^\alpha} = Nq^m L_\lambda(p,m)/(\zeta(\lambda,m)),
\]

where \( L_\lambda(p,m) \) is the generalized polylogarithm function, and \( L_\lambda(1,m) = \zeta(\lambda,m) \) [23]. Equation (1) is valid for \( \lambda > 1 \).

As \( q \to 1 \) (or \( p \to 0 \)), we have \( n_s \to N \). Thus for a fixed network size \( N \), there exists a critical value \( q_c \) of the prob-
ability $q$ 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 $\lambda$ cases where $L_i(p,m) / \xi(\lambda,m)$ tends to one, thus the number of blind spots becomes $n_s \sim Nq^{p_m} \leq Np^m$. We then have

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

where $\alpha = 1/m$ is the algebraic scaling exponent. In general, there is no closed form of $L_i(p,m)$, so a precise scaling law cannot be obtained explicitly. However, since $L_i(p,m) = \frac{1}{m} \cdot \frac{2}{1+m} \cdots = m^{a-1}(1+\left(\frac{1}{1+m}\right)^p)$, 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 $(1+\left(\frac{1}{1+m}\right)^p)$. This correction diminishes for large $\lambda$. A linear fit of the data in Fig. 4(b) gives the numerical value of the algebraic scaling exponent as $\alpha = 0.1$, 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 $q_b$ is far beyond the percolation threshold $q_b$, we can neglect the higher order blind spots. The number of blind spots is then $n_s = Np_b^{m}L_i(p_b,m) / \xi(\lambda,m)$. 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 $k \neq k_1$ or $k \neq k_m$ for $k_1 \leq k_m$, where $k_1$ and $k_m$ are the minimal and maximal degree of the network, respectively. The values of $k_1$ and $k_m$ are chosen to optimize the network's robustness to random node failures for a given average degree $k$ [11,12]. A similar calculation yields $n_s = Nq^{k_m}(k_m - k_1)(k_m - k_2) + p_{bc}(k_m - k_1)$. The inset of Fig. 2 shows the simulation result (symbols) and the theory (curves) of two cases with $\langle k \rangle = 20$: $k_1 = 10, k_m = 30$ (squares) and $k_1 = 15, k_m = 25$ (circles). Since the range $[k_1, k_m]$ 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.


