Resilience of complex networks to random breakdown

Gerald Paul,* Sameet Sreenivasan, and H. Eugene Stanley

Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA
(Received 1 July 2005; published 23 November 2005)

Using Monte Carlo simulations we calculate f_c, the fraction of nodes that are randomly removed before global connectivity is lost, for networks with scale-free and bimodal degree distributions. Our results differ from the results predicted by an equation for f_c proposed by Cohen *et al.* We discuss the reasons for this disagreement and clarify the domain for which the proposed equation is valid.

PACS number(s): 84.35.+i, 02.50.Cw, 05.50.+q, 64.60.Ak

I. INTRODUCTION

Recently, there has been much interest in the resilience of real-world networks to failure of network nodes [1–9]. It has been shown [3,4,6,7] that random uncorrelated networks with degree distribution $P(k)$ lose global connectivity when

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} < 2. \tag{1}$$

For nonrandom (degree correlated) networks there are two equivalent criteria for determining when global connectivity is lost.

(i) Global connectivity is lost when [8]

$$\det(\mathbf{A}) = 0 \tag{2}$$

where $\mathbf{A}$ is a matrix containing elements $A_{j,k} = k e_{j,k} + q_j \delta_{i,j}$ with $e_{j,k}$ the joint probability of the remaining degrees [10] of the two vertices at either end of a randomly chosen edge and with q_k the probability of the remaining degree of a single vertex at the end of a randomly chosen edge.

(ii) Equivalently, global connectivity is lost when [9] the maximum eigenvalue of the matrix $(k'-1)P(k'|k)$ is larger than 1 where $P(k'|k)$ is the probability that a node of degree k is connected to a node of degree k'.

At present, for correlated networks there is no closed form expression which specifies the value of the critical threshold f_c given the degree distribution and correlations of a given network. However for random networks such an expression has been derived [3,6]. As explained in Refs. [3,6], random removal of a fraction f of nodes from a network with degree distribution $P_0(k)$ results in a new degree distribution

$$P(k) = \sum_{k_0=k}^{K} P_0(k_0) \binom{k_0}{k} (1-f)^k f^{k_0-k}. \tag{3}$$

Using this degree distribution to calculate $\langle k \rangle$ and $\langle k^2 \rangle$ after random removal of sites it was determined [3,6] that

$$f_c = 1 - \frac{1}{\kappa_0 - 1} \tag{4}$$

where κ_0 is the value of κ computed from the original degree distribution, before the random removal. Equation (4) was

observed to hold for a number of network types, including random networks that have a Poisson degree distribution, and was used in the analysis of scale-free networks that have power-law degree distributions [3,6].

Using Monte Carlo simulations we find that Eq. (4) does not hold for networks with (i) self-loops and multiple edges and/or (ii) high variance in f_c. We illustrate our findings using scale-free and bimodal networks and clarify the domains where Eq. (4) is valid.

II. MONTE CARLO SIMULATIONS

We create random networks having specified degree distributions using the method described in Ref. [7]. We then randomly delete nodes in the network and after each node is removed, we calculate κ. When κ becomes less than 2 we record the number of nodes i removed up to that point [11]. This process is performed for many realizations of random graphs with a specified degree distribution and, for each graph, for many different realizations of the sequence of random node removals. The threshold f_c is defined as

$$f_c \equiv \frac{\langle i \rangle}{N} \tag{5}$$

where $\langle i \rangle$ is the average value of i.

III. SCALE-FREE NETWORKS

We study scale-free random networks with degree distribution

$$P(k) \sim k^{-\lambda} \quad (m \leq k \leq K). \tag{6}$$

We choose the lower cutoff $m=4$ and the upper cutoff $K=N$. In Figs. 1(a)–1(c) for $N=10^2$, 10^4, and 10^6 [12] we show the dependence on λ of $1-f_c^{MC}$ obtained by the Monte Carlo simulations and compare it with $1-f_c^{th}$ obtained theoretically from Eq. (4). The simulation results agree well with Eq. (4) for $\lambda > \lambda^*$, where $\lambda^* \approx 3$, and the agreement becomes better for increasing N. However, for $\lambda < \lambda^*$ there is significant disagreement, and the disagreement becomes larger as N increases, as seen clearly Fig. 1(d) in which we plot the normalized difference

$$\Delta \equiv \frac{f_c^{th} - f_c^{MC}}{f_c^{MC}}. \tag{7}$$
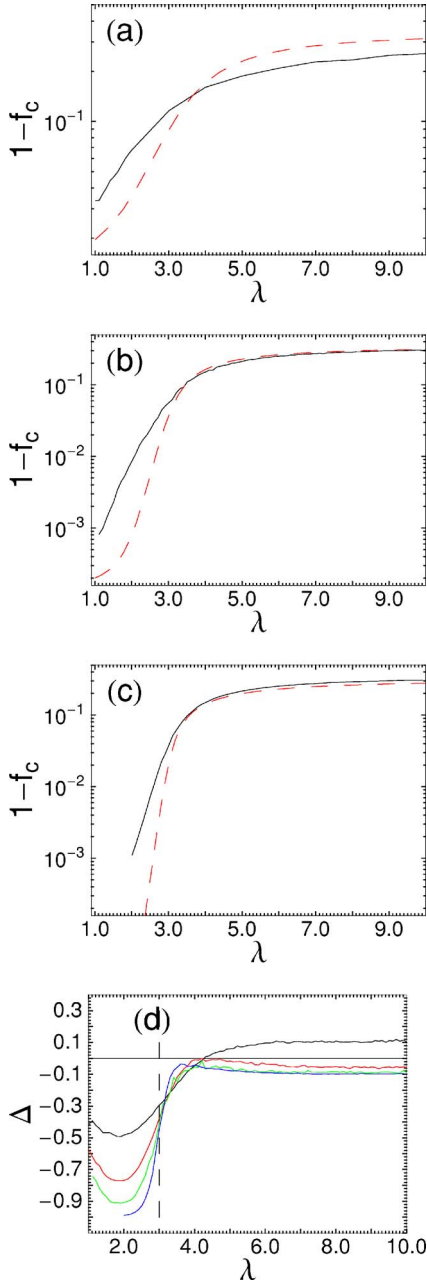
---

*Electronic address: gerryp@bu.edu

FIG. 1. (Color online) For $N=10^2$, 10^4, and 10^6, respectively, in (a), (b), and (c), $1-f_c$ versus λ. The solid line represents the results of Monte Carlo simulations; the dashed line is the prediction of Eq. (4). (d) The difference Δ [see Eq. (7)] between the prediction of Eq. (4) and Monte Carlo simulations for (from top to bottom) $N=10^2, 10^3, 10^4, 10^6$. Note that if we had used a larger value of the upper cutoff K, then Δ would decrease monotonically from $\lambda=3$ to 1 instead of having a minimum near $\lambda=2$.

The nonzero value of Δ has its root in the use of Eq. (3) to derive Eq. (4). Equation (3) is valid only if, in the original network, two conditions hold: (i) There are no *self-loops*, i.e., all links from node i are to distinct nodes j with $j \neq i$, and (ii) there are no multiple links between i and j. In graph theory networks satisfying these two conditions are called *simple*. If the original network is not simple, Eq. (3) must then be interpreted as operating on the original network but with
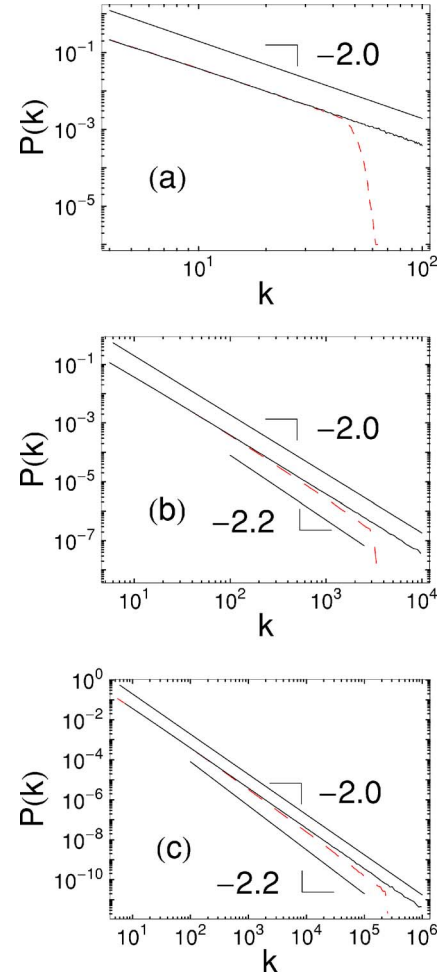

FIG. 2. (Color online) $P(k)$ versus k for $N=10^2$, 10^4, and 10^6 in (a), (b), and (c), respectively. The solid line represents $P(k)$ after network construction using the Molloy-Reed method; the dashed line is the distribution after the removal of self-loops and multiple edges.

self-loops and multiple links deleted. But this deletion changes the properties of the degree distribution. As seen in Figs. 2(a)–2(c) the cutoff is changed, and for large N, the slope of the tail of the distribution is modified. Also the degrees of adjacent nodes become correlated as seen in Fig. 3, which shows the λ dependence of the degree correlation [8]

$$r \equiv \frac{1}{\sigma_q^2} \sum_{j,k} (e_{jk} - q_j q_k), \qquad (8)$$

where e_{jk} is the joint probability of the remaining degrees of the two vertices at either end of a randomly chosen edge, q_k is the probability of the remaining degree of a single vertex at the end of a randomly chosen edge, and

$$\sigma_q^2 \equiv \sum k^2 q_k - \left(\sum_k k q_k \right)^2. \qquad (9)$$

Because of the degree correlations, Eq. (1) no longer applies and therefore Eq. (4) no longer holds. The similarity in appearance between Fig. 1(d) and Fig. 3 confirms that the non-
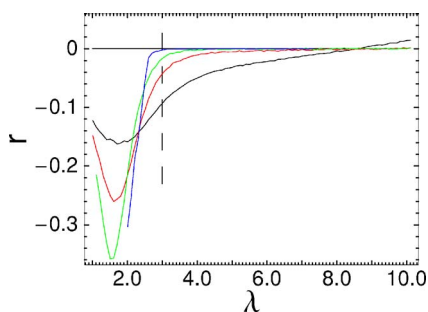
FIG. 3. (Color online) Correlation r as a function of λ for (from top to bottom at left) $N=10^2$, 10^3, 10^4, and 10^6 for distributions after removal of self-loops and multiple edges. Note that the correlation increases with N for $\lambda \lesssim 3$ and decreases with N for $\lambda \gtrsim 3$.

zero correlations play a major role in the difference between f_c^{MC} and f_c^{th}.

We can explain the domain of validity of Eq. (4) as follows. It is known [13–16] that for any desired random degree distribution, the networks created by such methods as those of Molloy and Reed [7] or Chung and Lu [13] create simple graphs only if $P(k)=0$ for k greater than the *structural cutoff*

$$K_s \equiv \sqrt{\langle k \rangle N}. \tag{10}$$

It is also known that for scale-free networks the number of nodes with degree greater than the *natural cutoff*

$$K_c \equiv mN^{1/(\lambda-1)} \tag{11}$$

is statistically insignificant [3,17]. These two facts are sufficient to understand that Eq. (4) is valid for scale-free networks only if $\lambda>3$ (in which case the natural cutoff K_c results in nodes with degree $\gtrsim \sqrt{N}$ being statistically insignificant) or for $\lambda<3$ if the maximum degree is less than the structural cutoff K_s.

IV. BIMODAL NETWORKS

A. Star networks

First, we discuss a simple example with a bimodal degree distribution for which Eq. (4) fails. Consider a *star network* of N nodes with degree distribution

$$P(k) = \begin{cases} (N-1)/N & (k=1), \\ 1/N & (k=N-1) \end{cases} \tag{12}$$

and $P(k)=0$ for all other values of k. If nodes are randomly removed, the criterion for losing global connectivity, $\kappa<2$, is obtained when the single node with degree $N-1$, the *hub* node, is removed or when almost all of the degree 1 nodes, the *leaf* nodes, are removed. The probability that almost all the leaf nodes are removed before the hub node is removed approaches 0 for large N. Let i be the number of nodes that are removed before the hub node is removed. Since the removal is random, i is uniformly distributed between 0 and $N-1$ and, from Eq. (5), $f_c=1/2$. On the other hand, Eq. (4) predicts $f_c=1-2/N$ which asymptotically approaches unity for large N.

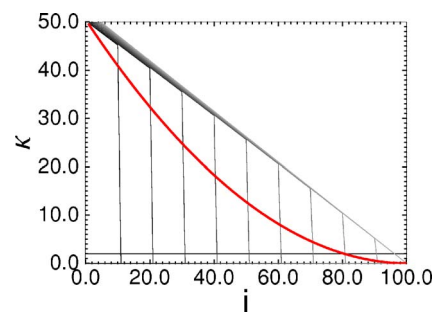As for the case of scale-free networks, we can understand this disagreement as a result of the presence of self-loops.

FIG. 4. (Color online) Example illustrating case in which $\langle (i|\kappa=2) \rangle \neq (i|\langle \kappa \rangle=2)$ for star network of one hub of degree 99 and 99 nodes of degree 1. Thin lines are κ vs i, where i denotes the number of the step at which a node is deleted, for cases in which the hub is deleted at step (from left to right) 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. The thick line is the average of the thin lines. Note that the value of i at which the average crosses the horizontal line $\kappa=2$ is much higher than 50, the average of the values of i at which the thin lines cross the horizontal line $\kappa=2$.

We can also use this star network example to identify another implicit assumption used in the derivation of Eq. (4), namely, that

$$\langle i \rangle \equiv \langle (i|\kappa(i)=2) \rangle = (i|\langle \kappa(i) \rangle = 2) \tag{13}$$

where $\kappa(i)$ is the value of κ after the removal of i nodes [18]. That is, we define $\langle i \rangle$ to be the average of i such that in each random removal $\kappa(i)=2$; the derivation of Eq. (4) assumes that $\langle i \rangle$ is equal to i such that the average of $\kappa(i)$ over all random removals equals 2. Equation (13) will be true in the limit in which the variance $\langle (i-\langle i \rangle)^2 \rangle$ is zero. But when the variance becomes large as is the case for the star network, Eq. (13) may not hold. Figure 4 illustrates graphically an example for which Eq. (13) does not hold because the variance in i is large.

B. Other bimodal networks

In order to study other bimodal networks, we extend the star network to networks with q high degree hubs connected to the remaining nodes of degree 1. For networks with average degree $\langle k \rangle$, the degree distribution is specified as

$$P(k) = \begin{cases} (N-q)/N & (k=1), \\ q/N & (k=k_2), \end{cases} \tag{14}$$

where

$$k_2 = \frac{(\langle k \rangle - 1)N + q}{q}, \tag{15}$$

and $P(k)=0$ for all other k. We first consider networks with $\langle k \rangle=2$. In Fig. 5(a), for the distribution of Eqs. (14) and (15), we plot $1-f_c$ as a function of q for $N=10^2$, 10^3, 10^4, 10^5, and 10^6. Also shown in Fig. 5(a) are plots for approximations f_c^{high} and f_c^{low} which we expect to be valid, respectively, for high and low values of q. We will use these approximations to determine how $f_c(q)$ scales and for which values of q Eq. (4) is valid. The approximations are determined as follows.
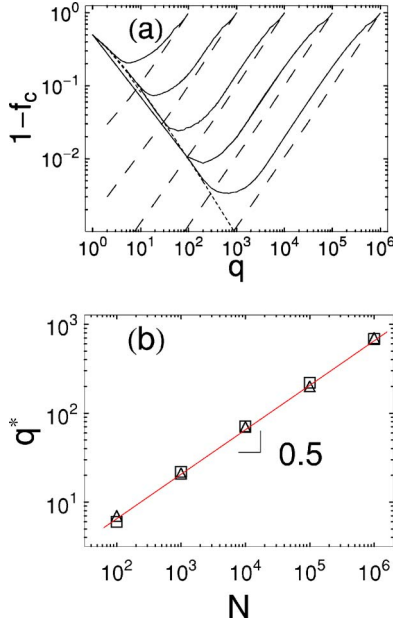
FIG. 6. (Color online) Number of hubs q^* at which approximations for low and high q intersect vs N. Squares, triangles, and circles represent networks with $\langle k \rangle = 2$, 3, and 4, respectively.

FIG. 5. (Color online) For $\langle k \rangle = 2$ and for (from left to right) $N = 10^2$, 10^3, 10^4, 10^5, and 10^6 (a) $1 - f_c$ vs number of hubs q. The solid lines represent Monte Carlo simulation results. Dashed lines (short) are the approximation f_c^{low}; dashed lines (long) are the approximation f_c^{high}. (b) Number of hubs q versus N. Squares represent characteristic values q^* at which high and low q approximations intersect. Triangles represent values of q at which the standard deviation in $1 - f_c$ is minimal.

(i) When $q \sim N$, (i.e., the network is homogeneous) we expect Eq. (4) to hold so $f_c^{\text{high}} = 1 - 1/(\kappa_0 - 1)$.

(ii) For small q, the network loses global connectivity when all q high degree nodes are removed. The probability that all q high degree nodes are removed after the first i nodes of all types have been removed is

$$g(q,N,i) = \frac{q}{N} \frac{\binom{i-1}{q-1}}{\binom{N-1}{q-1}}. \tag{16}$$

Here i is now the average number of nodes that must be removed before all q high degree nodes are removed. Then $\langle i \rangle = \sum_{i=q}^{N} i g(q,N,i)$ and

$$f_c^{\text{low}} = \frac{\langle i \rangle}{N} = \frac{\sum_{i=q}^{N} i g(q,N,i)}{N}. \tag{17}$$

Note that f_c^{low} does not depend on $\langle k \rangle$ since changing $\langle k \rangle$ results simply in a different number of links between the high degree nodes; if our criterion for collapse is the removal of all high degree nodes, the number of links between them is irrelevant. As expected, the plots of f_c^{low} and f_c^{high} approximate the values of f_c for low and high values of q, respectively.

In Fig. 5(b), we plot the number of hubs q^* for which the functions $f_c^{\text{low}}(q)$ and $f_c^{\text{high}}(q)$ intersect. We find that
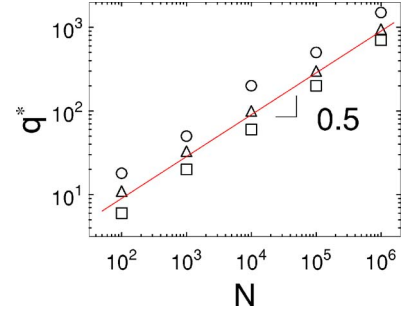
$$q^* \sim N^{0.5}. \tag{18}$$

Similar plots (see Fig. 6) for $\langle k \rangle = 3$ and 4 also exhibit scaling of q^* as $N^{0.5}$ with only a change in the prefactor; the scaling is independent of $\langle k \rangle$.

The simulation results suggest that q^* scales as $\sqrt{N}$. We can show this to be the case by solving analytically for q^* for large N as follows: For general $\langle k \rangle$, using the distribution in Eqs. (14), we find for $N \gg q \gg 1$

$$f_c^{\text{high}} = 1 - \frac{q}{(\langle k \rangle - 1)N}. \tag{19}$$

For f_c^{low}, the sum in Eq. (17) can be performed analytically, yielding

$$f_c^{\text{low}} = \frac{\Gamma(N+2)[\Gamma(q+2) - \Gamma(q+1)]}{N\Gamma(N+1)\Gamma(q+2)} \tag{20}$$

for $q > 0$. For large N,

$$f_c^{\text{low}} = \frac{\Gamma(q+2) - \Gamma(q+1)}{\Gamma(q+2)}. \tag{21}$$

To first order in $1/q$, Eq. (21) yields

$$f_c^{\text{low}} = 1 - \frac{1}{q} + O\left(\frac{1}{q^2}\right). \tag{22}$$

Equating Eqs. (22) and (19) we find

$$q^* = \sqrt{\langle k \rangle - 1}\sqrt{N} \tag{23}$$

consistent with the plot in Fig. 5(b) and Eq. (18). From the fact that q^* scales like $\sqrt{N}$, we conclude that all characteristic values of f_c scale like $\sqrt{N}$ with a prefactor dependent on $\langle k \rangle$. In particular the value of q at which f_c^{MC} (found from Monte Carlo simulations) agrees to any desired degree with the value of f_c^{th} [from Eq. (4)] will scale with N in the same fashion in which q^* scales with N, Eq. (18). For simplicity, we consider Eq. (3) to be valid for $q > q^*$.

We now confirm that the variance in f_c is in fact small for values of q for which Eq. (13) holds. In Fig. 7(a), for $N = 10^3$ and $q = 1$, 5, 10, and 20, we plot $P(1 - f_c)$ vs $1 - f_c$. As expected, for $q = 1$ (star network) the distribution is uniform because there is an equal probability that the single high degree node will be removed at any value of i. For the larger
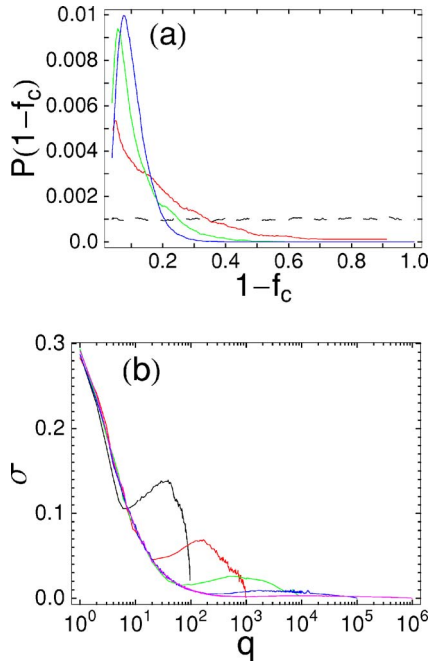
FIG. 7. (Color online) (a) $P(1-f_c)$, the probability distribution of $1-f_c$ for $N=10^3$ and $q=1$ (dashed line) and (from left to right in order of increasing position of peaks) $q=5, 10$, and 20. (b) Standard deviation σ versus q for $N=10^2, 10^3, 10^4, 10^5$, and 10^6 (from left to right in order of increasing length of the tails of the distributions). Note that the second peak in this plot which is most pronounced for smaller N is an artifact of finite size.

values of q, the distributions $P(1-f_c)$ develop a well-defined peak. To quantify the definition of these peaks, we plot in Fig. 7(b), the standard deviation of f_c,

$$\sigma = \frac{\sqrt{\langle i^2 \rangle - \langle i \rangle^2}}{N}, \tag{24}$$

versus q for $N=10^2, 10^3, 10^4, 10^5$, and 10^6. Each of the plots has a large deviation at $q=1$ and decreases to a local minimum, the position of which, $\tilde{q}$, increases with increasing N. For q greater than $\tilde{q}$, the deviation is small and decreases with increasing N. In Fig. 5(b) we plot $\tilde{q}$ as a function of N. We see that the values of these minima are essentially the same as the values of q^*, the value of q above which Eq. (4) is valid. This is consistent with our understanding that Eq. (4) is valid when the variance is small.

C. Domain of validity

Since q and the degree of the hubs k_2 are related by Eq. (15), we can determine for what values of k_2 Eq. (13) is valid. Substituting Eq. (23) in Eq. (15) we find that Eq. (13) is valid when
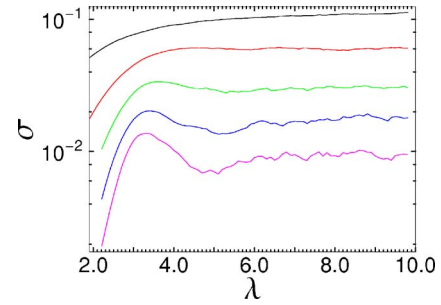

FIG. 8. (Color online) For random scale-free networks with $4 \leq k \leq N$, standard deviation σ_{f_c} versus λ for $N=10^2, 10^3, 10^4, 10^5$, and 10^6 (from top to bottom).

$$k_2 < \sqrt{(\langle k \rangle - 1)N}. \tag{25}$$

Thus the criterion for Eq. (13) holding is essentially the same as the criterion discussed in Sec. III for the graph being simple. The bimodal networks we study here in which a relatively small number of nodes control the global connectivity of the network yield large variances in f_c for networks with a given number of nodes; in fact, for $q=1$ the worst case variance is obtained. This suggests that the criterion of Eq. (25) may hold for all degree distributions as a requirement for a low variance in f_c. If this is the case, we can use the requirement that $P(k)=0$ for $k \lesssim K_c$ as the criterion for both the network being simple and f_c having a small variance. Note, however, that while the criteria are similar, it is not true that the presence of self-loops and multiple edges implies that the distribution of f_c has a large variance; for example, the variance of f_c in scale-free networks is small even in the presence of self-loops and multiple edges, as seen in Fig. 8.

V. DISCUSSION AND SUMMARY

We have clarified the domain of validity of Eq. (4), a general equation for determining f_c, the fraction of nodes that must be randomly removed before global connectivity is lost. For Eq. (4) to be valid, (i) the highest degree of any nodes present in statistically significant numbers in a random network must be less than the structural cutoff $K_s \equiv \sqrt{\langle k \rangle N}$ and (ii) the variance of f_c must be small. For bimodal networks the variance in f_c is small when the hubs have degree less than $\sqrt{(\langle k \rangle - 1)N}$. That the bimodal networks we have studied represent a worst case for large variance suggests that in general the criterion that the network be simple is sufficient for Eq. (13) to hold. It is not clear if there is a deeper connection between these two criteria, both of which scale as $\sqrt{N}$.

[1] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).

[2] V. Paxon, IEEE/ACM Trans. Netw. **5**, 601 (1997).

[3] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[4] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[5] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[6] R. Cohen, D. ben-Avraham, and S. Havlin, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, New York, 2002).

[7] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[8] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[9] A. Vázques and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).

[10] The *remaining degree* of a vertex is defined as the degree of the vertex -1.

[11] Note that we perform all simulations on networks created by the Molloy-Reed method keeping all self-loops or multiple edges created in the process. Thus the network is random and has no degree correlations and since we include the contribution of self-loops and multiple edges in the calculation of κ the criterion $\kappa < 2$ is valid.

[12] The plot for $N = 10^6$ is limited to $\lambda > 2$. For $\lambda < 2$ the number of edges in the network increases rapidly and exceeds available computer memory.

[13] F. Chung and L. Lu, Ann. Comb. **6**, 125 (2002).

[14] Z. Burda and A. Krzywicki, Phys. Rev. E **67**, 046118 (2003).

[15] M. Boguñá, R. Pastor-Sartorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).

[16] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).

[17] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[18] The notation $(x|y)$ should be read as "x such that condition y holds."