Probabilistic Prediction in Scale-Free Networks: Diameter Changes

J.-H. Kim, K.-I. Goh, B. Kahng, and D. Kim

School of Physics and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea

(Received 5 December 2002; published 1 August 2003)

In complex systems, responses to small perturbations are too diverse to definitely predict how much they would be, and then such diverse responses can be predicted in a probabilistic way. Here we study such a problem in scale-free networks, for example, the diameter changes by the deletion of a single vertex for various *in silico* and real-world scale-free networks. We find that the diameter changes are indeed diverse and their distribution exhibits an algebraic decay with an exponent ζ asymptotically. Interestingly, the exponent ζ is robust as $\zeta \simeq 2.2(1)$ for most scale-free networks and insensitive to the degree exponents γ as long as $2 < \gamma \leq 3$. However, there is another type with $\zeta \simeq 1.7(1)$ and its examples include the Internet and its related *in silico* model.

DOI: 10.1103/PhysRevLett.91.058701 PACS numbers: 89.75.Hc, 05.10.–a, 89.20.Hh, 89.75.Da

A complex system consists of many constituents, generating emerging behavior through diverse interactions [1,2]. One of the powerful ways of examining the intrinsic nature of a complex system is to observe how such emerging patterns change by the small perturbation applied to the system. In complex systems, such a change or response is so sensitive to the details of the perturbation that it is extremely diverse. In such a case, it is not adequate to *predict* definitely how much the change would be. Recently, Parisi argued [3] that the prediction for the responses to small perturbations in complex systems can be made in a probabilistic way. He showed examples of protein structures in biological systems and spin glasses in physical systems. In the case of proteins, subject to small external perturbations such as the change in *p*H or the substitution of a single amino acid, they would fold to a completely different 3D structure but with practically the same free energy. In the case of the disordered magnetic systems, each spin responds to a slowly varying external field by changing its orientation, forming a series of bursts, known as Barkhausen noise [4]. The number of spin bursts depends on the disorder strength of the system, following a power-law distribution at a critical strength of disorder. The prediction of the number of spin bursts in this case can only be probabilistic. The stock market is another example of complex systems. Stock prices are determined as a result of the complicated interplay between numerous investors, and the price changes were also found to exhibit a power-law distribution [5]. All these examples aptly illustrate how the concept of probabilistic prediction may apply as a new paradigm in modern science. Other examples can also be found in fields as diverse as meteorology and geology [6].

Recently, there have been many works which describe complex systems in terms of graphs [7,8], where vertices represent constituents and edges interactions between constituents. An interesting feature emerging in such complex networks is the emergence of a power-law behavior in the degree distribution, $P_{\mathrm{d}}(k) \sim k^{-\gamma}$ [9], where the degree k is the number of edges incident upon a given vertex. Such complex networks are called scale-free (SF) networks.

In this Letter, we study how SF networks respond to small perturbations and check if the concept of probabilistic prediction can be applied. For this purpose, we investigate a simple problem of diameter change when a single vertex is removed from the system. Diameter, defined as the average distance between every pair of vertices in a network, is a simple yet fundamental quantity of SF networks to characterize the small-world nature and can be thought of as a measure reflecting the efficiency of a network. Our main interest is how much the efficiency of a network would be affected by the removal of a single vertex. When a vertex is removed, each pair of remaining vertices whose shortest pathway had passed through the removed vertex should find detours, resulting in the rearrangement of shortest pathways over the network. Thus, the diameter change occurs in a collective manner. From extensive numerical calculations for a number of SF network models and real-world examples, we find that the diameter changes indeed are very diverse and crucially depend on the degree of the removed vertex. When a vertex with a small number of connections is removed, the diameter changes little. However, when a vertex with a large number of connections is removed, the diameter change is drastic, exhibiting a power-law distribution with an exponent ζ,

$$P_{\mathrm{c}}(\Delta) \sim \Delta^{-\zeta}, \tag{1}$$

for large Δ, where Δ is the dimensionless relative diameter change defined as the diameter change caused by the removal of a certain vertex divided by the original diameter before the removal, and $P_{\mathrm{c}}(\Delta)$ is its distribution. Moreover, the exponent ζ turns out to be robust for various SF networks, insensitive to the degree exponent γ for $2 < \gamma \leq 3$.

To be specific, we consider an undirected SF network with a finite number of vertices N and measure the diameter of the network. Note that we limit our interest

to undirected networks only in this work. Next, we remove a certain vertex i and measure the diameter d_i of the rest of the network. Measuring a dimensionless quantity $\Delta_i = (d_i - d_0)/d_0$ for all i, where d_0 is the diameter of the original unperturbed network, we obtain the distribution of Δ for the network. Note that our case is different from the previous study of the robustness of SF networks [10,11] where vertices are removed successively. In our case, on the other hand, only a single vertex is removed each time. When a certain vertex is removed, the network may disintegrate into more than one cluster. In such cases, d_i is calculated only within the largest cluster. The diameter can be measured via a simple breadth-first search algorithm. To obtain the distribution of the diameter changes, we need computation time of order $\mathcal{O}(N^3)$.

To look into more details, we consider the static model [12]. It is constructed by connecting mN pairs of vertices i and j with probability proportional to $(ij)^{-\alpha}$, where N is the vertex number and α is a parameter. We use $m = 2$. Its degree distribution follows a power law $P_d(k) \sim k^{-\gamma}$ with $\gamma = 1 + 1/\alpha$. Thus, tuning the parameter α in $[0, 1)$, we obtain a continuous spectrum of the exponent γ in the range $2 < \gamma < \infty$.

The diameter change Δ_i by the removal of a certain vertex i in such SF networks can be positive or negative, and the histogram of the diameter changes is highly centralized around $\Delta = 0$ (Fig. 1). However, it exhibits a fat tail for $\Delta > 0$ (the inset in Fig. 1). For the static model with $\gamma = 3$, for example, the case of small diameter changes in the range $|\Delta| < 2 \times 10^{-4}$ occurs with frequency as high as 96%. Thus, the effect of a vertex removal usually is negligible as a whole, which is manifested by the exponentially bounded fluctuations of the diameter around its original value. We estimate the N dependence of such small diameter changes in a mean-field-type approach. It is known that the diameter d_0 depends on the number of vertices as $d_0 \sim \ln N$ for a random graph and $d_0 \sim \ln N/\ln\ln N$ [13] for the Barabási-Albert model [9] with $\gamma = 3$. When a vertex is removed, the diameter may be reduced as $d \sim \ln(N-1)$ or $d \sim \ln(N-1)/\ln(N-1)$, both leading to $\Delta \approx -1/N\ln N$ for large N. Thus, when $N = 10^4$, $\Delta \sim \mathcal{O}(10^{-5})$, which is comparable to numerical values of the central part in Fig. 1. On the other hand, substantial (about 4%) vertices have a serious impact on the system's efficiency and they indeed contribute to the positive tail of the histogram, showing the power-law behavior, Eq. (1). We find that such large diameter changes are due mainly to the removal of a vertex with large degree. This feature is reminiscent of the percolation problems on the SF networks [14,15].

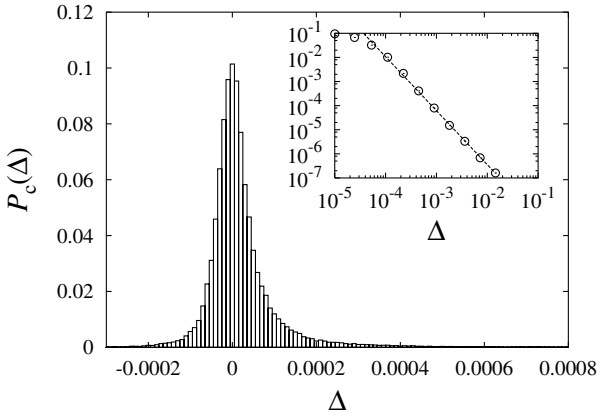

FIG. 1. Normalized histogram of the diameter changes for the static model with $\gamma = 3$ and $N = 10^4$, averaged over ten configurations. Horizontal range is truncated for clearance, but runs up to 2×10^{-2}. Inset: Plot of $P_c(\Delta)$ in log-log scale for $\Delta > 0$. The dashed line is a fit line having a slope -2.2. Data points are logarithmically binned.

Let us investigate the power-law behavior for large Δ in details. The exponent ζ seems to be robust as $\zeta \approx 2.2(1)$ as long as $2 < \gamma \leq 3$ for the static model as shown in Fig. 2. Similar behaviors are found in other model networks (ii)–(vii) listed in Table I. These include the SF networks showing nontrivial degree-degree correlations [26]. For $\gamma > 3$, on the other hand, as γ increases, the power-law behavior sets in only for larger values of Δ and the exponent ζ increases with γ. Eventually, the diameter change distribution for the Erdös-Rényi random networks decays exponentially as shown in Fig. 2.

To see such universal behavior of ζ in the real world, we consider a couple of real-world networks, the protein interaction networks (PIN) and the Internet. For the PIN of the yeast *Saccharomyces cerevisiae* [23], we also find a power law in the diameter change distribution with an exponent $\zeta \approx 2.3(1)$ (Fig. 3), consistent with the one obtained for various model networks, including the one proposed as its own *in silico* model (vii) [22].

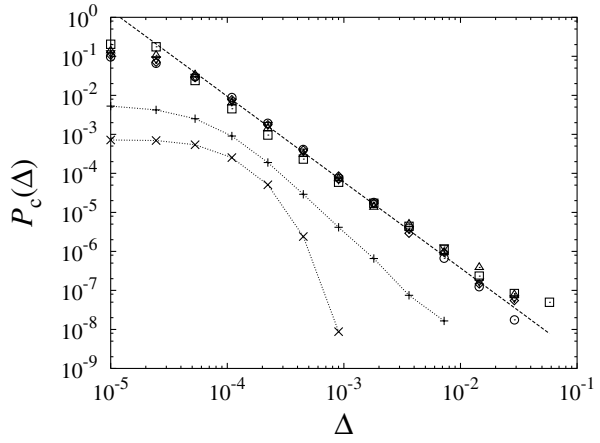

FIG. 2. The diameter change distribution $P_c(\Delta)$ for the static model with $\gamma = 2.2$ (□), 2.4 (◇), 2.6 (▽), 2.8 (△), 3.0 (○), and 4.0 (+), and the Erdös-Rényi model (×). The data, obtained for $N = 10^4$ and averaged over ten configurations. The two data sets (+, ×) are shifted vertically for comparison. The dashed line having a slope -2.2 is drawn for the eye. Note that the deviations from the straight line at the fat tail are due to the generic finite-size effects for the SF networks with $\gamma < 3$ [16].

TABLE I. Summary of the results for various SF networks. Tabulated for each network are the system size N, the mean degree $\langle k \rangle$, the degree exponent γ, the diameter change exponent ζ, and the betweenness centrality exponent η [17].

System	N	$\langle k \rangle$	γ	ζ	η	Ref.
(i) Static model	10^4	4	2.2–3.0	2.2(1)	2.2(1)	[12]
(ii) Barabási-Albert model	10^4	4	2.2–3.0	2.2(1)	2.2(1)	[9]
(iii) Copying model	10^4	4	2.2–3.0	2.2(1)	2.2(1)	[18]
(iv) Fitness model	10^4	4	2.25	2.2(1)	2.2(1)	[19]
(v) Accelerated-growth model	10^4	$\mathcal{O}(1)$	3.0(1)	2.2(1)	2.2(1)	[20]
(vi) Huberman-Adamic model	10^4	$\mathcal{O}(1)$	3.0(1)	2.2(1)	2.2(1)	[21]
(vii) Protein interaction network model	10^4	$\mathcal{O}(1)$	⋯	2.2(1)	2.2(1)	[22]
(viii) Protein interaction network of the yeast *S. cerevisiae*	5662	6.1	3.2(2)	2.3(1)	2.3(1)	[23]
(ix) Internet at the autonomous systems level	6474, 12 058	~4	2.1(1)	1.7(1)	2.0(1)	[24]
(x) Adaptation model	~6500	$\mathcal{O}(1)$	2.1	1.7(1)	2.0(1)	[25]

For the Internet at the autonomous systems level [24], the diameter change distribution again follows a power law, however, with a different exponent $\zeta \approx 1.7(1)$ (Fig. 4). The smaller exponent ζ indicates that the effect of the removal of vertices contributing to the tail of the distribution is much more severe than the previous cases with $\zeta \approx 2.2$ [(i)–(viii) in Table I]. To confirm the novel value of ζ for the Internet, we perform the same calculations for its *in silico* model, called the adaptation model [25], and indeed obtain $\zeta \approx 1.7$ for it, too. The two different behaviors of the diameter change distribution are rooted from distinct topological features of shortest pathways of each case, which will be discussed later.

Recently, it was proposed that the SF networks with $2 < \gamma \leq 3$ can be classified into two classes [12,17], following the power-law behavior of the betweenness centrality (BC) distribution [27,28]. The BC g_k of a vertex k is the accumulated sum of the fraction of shortest pathways passing through k and its distribution follows a power law, $P_g(g) \sim g^{-\eta}$ for SF networks. The BC exponent η turns out to be robust as either $\eta \simeq 2.2(1)$ (class I) or $\eta \approx 2.0(1)$ (class II) as long as $2 < \gamma \leq 3$ [12,17]. Interestingly, the networks (i)–(viii) in Table I having the diameter change exponent $\zeta \approx 2.2$ belong to class I, and the values of ζ and η coincide with each other within our numerical resolutions, while they are different for class II. Empirically, the rank of a vertex in g and that in Δ are likely to be the same for vertices with large degrees. If then, the relation $P_g(g)dg \sim P_c(\Delta)d\Delta$ would hold asymptotically, leading to

$$\Delta(g) \sim g^{(\eta-1)/(\zeta-1)} \tag{2}$$

for large g. This type of relation also holds between degree and BC [12]. Indeed, the slopes in the double

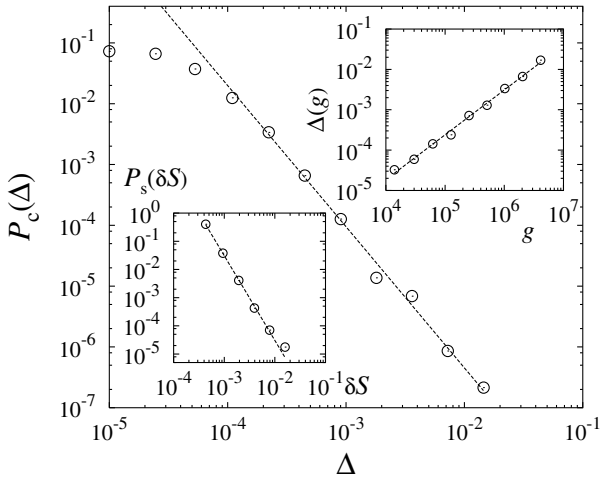

FIG. 3. The diameter change distribution $P_c(\Delta)$ for the PIN of the yeast *S. cerevisiae*. The slope of the fit line (dashed) is -2.3, drawn for the eye. Upper inset: Plot of $\Delta(g)$ vs. g. The slope of the straight line is 1.1, drawn for the eye. Lower inset: The largest-cluster-size change distribution $P_s(\delta S)$. Here δS is normalized by N. The slope of the fit line is -3.0, drawn for the eye.

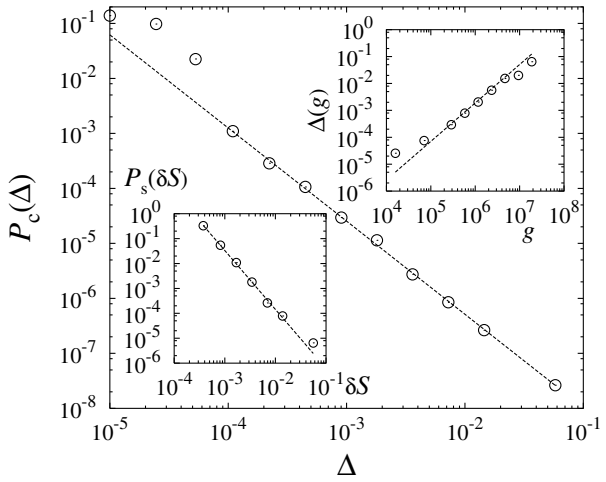

FIG. 4. The diameter change distribution $P_c(\Delta)$ for the Internet at the autonomous system level. The slope of the fit line (dashed) is -1.7, drawn for the eye. Upper inset: Plot of $\Delta(g)$ vs. g. The slope of the straight line is 1.4, drawn for the eye. Lower inset: The largest-cluster-size change distribution $P_s(\delta S)$. Here δS is normalized by N. The slope of the fit line is -2.4, drawn for the eye.

logarithmic scale in the upper insets in Figs. 3 and 4 are 1.1(1) for the PIN and 1.4(1) for the Internet, respectively, consistent with the predictions from the formula, Eq. (2). Thus, the two classes, classes I and II, are also categorized by the diameter change distribution and the distinction between them can be observed more clearly through it.

Our finding that the diameter change distribution is also classified into classes I and II following those for BC distribution may be rooted from the fact that both quantities, diameter and BC, depend on universal features of *the shortest pathways* topology between a vertex pair in networks. When the sum rule [29], i.e., $\sum_k g_k \sim d$, is applied, one can see immediately that the diameter change distribution is the same as the total BC change distribution. On the other hand, the networks belonging to class II are more sparse and ramified than those in class I, so that the Internet is more fragile by the removal of a single vertex than the PIN. We compare the distribution of the size change δS of the largest cluster for the PIN and the Internet by a single vertex removal. As shown in the lower insets in Figs. 3 and 4 , the giant cluster in the Internet becomes much smaller than in the PIN. Thus, the number of vertex pairs connected after the removal becomes much smaller in the Internet than in the PIN. Consequently, the difference of the exponent ζ between the two classes appears much larger than that of the exponent η in class II. However, it is not clear how the power-law behavior in $P_c(\Delta)$ arises and what determines its exponent.

It would be interesting to generalize our study to the case of having more than one vertex removed. For simplicity, we consider the case of two-vertex removal, in particular, one is the hub and the other is an arbitrary vertex. Interestingly, we find that the diameter change distribution also exhibits a fat-tail behavior with the same exponent ζ. We cannot check if the fat-tail behavior still holds for more general cases due to the huge amount of computing time. Meanwhile, it has been studied that the SF network is robust against random failures. To show this, the diameter change due to removal of a finite fraction of vertices was measured by taking an average over a few samples, not over the whole ensemble. If the diameter change distribution after those removals still possesses a power-law distribution with $\zeta \leq 3$, as is the case here, then we could say that the average diameter change cannot reflect its intrinsic nature because its variance diverges. Thus, the diameter changes after random failures should also be described in a probabilistic way.

In summary, we have studied the diverse behavior in response to a small perturbation, a deletion of a single vertex in SF networks. The diameter change Δ by a removal of a vertex is very diverse, exhibiting a power-law distribution with an exponent ζ for large Δ. Moreover, the diameter change exponent ζ is robust as $\zeta \simeq 2.2$ for most SF networks with $2 < \gamma \leq 3$, or $\zeta \simeq 1.7$ for the Internet as an exception.

This work is supported by the KOSEF Grant No. R14-2002-059-01000-0 in the ABRL program.

[1] K. Ziemelis and L. Allen, Nature (London) **410**, 241 (2001), and following review articles on complex systems.
[2] R. Gallagher and T. Appenzeller, Science **284**, 87 (1999), and following viewpoint articles on complex systems.
[3] G. Parisi, Physica (Amsterdam) **263A**, 557 (1999).
[4] J. P. Sethna, K. A. Dahmen, and C. R. Myers, Nature (London) **410**, 242 (2001), and references therein.
[5] H. E. Stanley *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, Suppl. 1, 2561 (2002), and references therein.
[6] M. Buchanan, Nature (London) **419**, 787 (2002).
[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[8] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[10] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[11] Z. Liu, Y.-C. Lai, and N. Ye, Phys. Rev. E **66**, 036112 (2002).
[12] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001); see also G. Caldarelli *et al.*, *ibid.* **89**, 258702 (2002).
[13] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[14] D. S. Callaway *et al.*, Phys. Rev. Lett. **85**, 5468 (2000).
[15] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000); **86**, 3682 (2001).
[16] Z. Burda and A. Krzywicki, Phys. Rev. E **67**, 046118 (2003); P. L. Krapivsky and S. Redner, J. Phys. A **35**, 9517 (2002).
[17] K.-I. Goh *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 583 (2002).
[18] R. Kumar *et al.*, in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Los Alamitos, 2000), p. 57.
[19] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[20] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 025101(R) (2001).
[21] B. A. Huberman and L. A. Adamic, Nature (London) **401**, 131 (1999).
[22] R. Solé *et al.*, Adv. Complex Syst. **5**, 43 (2002).
[23] K.-I. Goh, B. Kahng, and D. Kim (unpublished).
[24] The National Laboratory for Applied Network Research sponsored by the National Science Foundation provides Internet routing related information based on border gateway protocol data (see http://moat.nlanr.net/).
[25] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **88**, 108701 (2002).
[26] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[27] L. C. Freeman, Sociometry **40**, 35 (1977).
[28] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
[29] K.-I. Goh, B. Kahng, and D. Kim, Physica (Amsterdam) **318A**, 72 (2003).