Classes of the shortest pathway structures in scale free networks

K.-I. Goh, E. Oh, C.-M. Ghim, B. Kahng[1], and D. Kim

School of Physics, Seoul National University, Seoul 151-747, Korea
kahng@phya.snu.ac.kr

Abstract. We study a problem of data packet transport between a pair of vertices on scale-free network, and introduce load of a vertex as the accumulated sum of a fraction of data packets traveling along the shortest pathways between every pair of vertices. It is found that the load distributions for many real-world networks follow a power law with an exponent δ which is close to either 2.2(1) (class I) or 2.0 (class II), insensitive to different values of the degree exponent γ in the range, $2 < \gamma < 3$. The classification of scale-free networks into the two classes may stem from the characteristics of the shortest pathways structures. While the shortest pathways between a pair of vertices are multiply connected in the class I, they are almost singly connected in the class II. Such distinct topological features of the shortest pathways produce different behaviors in diverse problems such as the distribution of diameter change by the removal of a single vertex. Finally, we remark that since the two numerical values of the load exponent are too close, it still remains an open question if they are really robust. Analytic solution resolving this controversial issue is needed.

1 Introduction

Complex systems comprise of many elements such as individuals, substrates, and companies in social, biological, and economic systems, respectively, showing cooperative phenomena between elements through diverse interactions and adaptations to the pattern they create [1,2]. Such complex systems may be described in terms of graphs, consisting of vertices and edges, where vertices (edges) represent elements (their interactions) [3–6]. This approach was initiated by Erdős and Rényi (ER) [7]. In their model, the number of vertices N is fixed and edges are connected among vertices randomly. They found that the entire system can be connected with a small fraction of edges, being of order $\mathcal{O}(1/N)$. While such random graph theory is a pioneering work, it cannot describe realistic networks recently observed. In graph theory, one of interesting quantities is degree, defined as the number of edges connecting to a certain vertex. The degree distribution of the ER network follows a Poisson distribution.

Recently, there were findings that the degree distribution of the world-wide web follows a power law,

$$P_D(k) \sim k^{-\gamma}, \tag{1}$$

where k means degree and γ is the degree exponent. The network displaying a power-law degree distribution is called scale-free (SF) network. SF networks [8] are ubiquitous in real-world networks such as the world-wide web (WWW) [9–11], the Internet [12–14], the citation network [15] and the author collaboration network of scientific papers [16,17], and the metabolic networks in biological organisms [18]. To illustrate the mechanism of SF network formation, Barabási and Albert (BA) [19–21] introduced an evolving network model where the number of vertices N increases linearly with time rather than fixed, and a newly introduced vertex is connected to m already existing vertices with probability proportional linearly to the degree of the selected vertex, called preferential attachment (PA) rule. Then the degree exponent follows a power law with the exponent $\gamma = 3$. A generalized version assigns the probability proportional to $k + m(a-1)$, $a(>0)$ being a tunable parameter. Then the degree exponent is $\gamma = 2 + a$ [21].

Recently, we introduced another type of SF network model, called the static model, in which the number of vertices N is fixed from the beginning, and vertices are indexed by an integer i $(i = 1, \ldots, N)$ [22]. We assign the weight $p_i = i^{-\alpha}$ to each vertex, where α is a control parameter in $[0,1)$. We select two different vertices (i,j) with probabilities equal to the normalized weights, $p_i/\sum_k p_k$ and $p_j/\sum_k p_k$, respectively, and add an edge between them unless one exists already. This process is repeated until mN edges are made in the system. Then the mean degree is $2m$. Since edges are connected to a vertex with frequency proportional to the weight of that vertex, the degree at that vertex is given as

$$\frac{k_i}{\sum_j k_j} \approx \frac{(1-\alpha)}{N^{1-\alpha} i^{\alpha}}, \tag{2}$$

where $\sum_j k_j = 2mN$. Then it follows that the degree distribution follows the power law, Eq. (1), where γ is given by

$$\gamma = (1+\alpha)/\alpha. \tag{3}$$

Thus, adjusting the parameter α in [0,1), we can obtain various values of the exponent γ in the range, $2 < \gamma < \infty$.

Emergence of a power-law in the degree distribution in complex networks is a self-organized pattern in complex systems. From the viewpoint of the theory of equilibrium critical phenomena, the power-law behavior implies the absence of a characteristic scale, implying that the network is in a critical state. However, the degree exponent depends on the details of network structure, varying mostly in the range $2 < \gamma \leq 3$. Thus it would be interesting to find a physical quantity showing universal feature, insensitive to the details of network structure. In Section 2, we will introduce a quantity, called load, which has been used in the study of transport phenomena of data packet on networks. Interestingly, the load distribution displays a power law and the load exponent is likely to be insensitive to the details of SF network

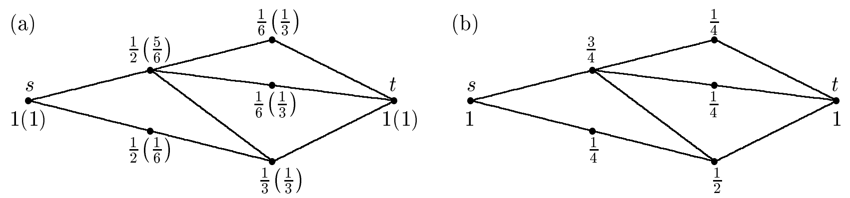

Fig. 1. The definitions of the load (a) and the BC (b). In (a), the load at each vertex due to a unit packet transfer from the vertex s to the vertex t, $\ell^{(s\to t)}$ is shown. In the parentheses, $\ell^{(t\to s)}$ is also shown for the packet transport from t to s for comparison. In (b), the BC at each vertex due to a packet transfer between s and t, $b^{(s,t)}$, is shown. The BC $b^{(s,t)}$ is equal to $b^{(t,s)}$.

structures. Currently it seems that there exist two different classes displaying different load exponents for real world networks. Furthermore, in Section 3, we will introduce the load-load correlation function, which behaves differently from the degree-degree correlation function in social networks. In Section 4, we will consider the distribution of diameter changes when a single vertex is removed. This problem is another example that can distinguish the shortest pathways characteristics between a pair of vertices in SF networks.

2 Load or betweenness centrality

Let us suppose that a data packet is sent from a vertex i to j, $(i \to j)$. It is transmitted along the shortest pathway between them. If there exist more than one shortest pathways, the data packet would encounter one or more branching points. In this case, the data packet is presumed to take one of them with equal probability, and the data packet is substantially divided evenly by the number of branches at each branching point as it travels. The contribution to the load at vertex k due to the packet flow from i to j is denoted as $\ell_k^{(i\to j)}$ as shown in Fig. 1a. Then we define the load ℓ_k of a vertex k, $\ell_k = \sum_{i,j} \ell_k^{(i\to j)}$. Note that $\ell_k^{(i\to j)}$ may be different from $\ell_k^{(j\to i)}$ even for undirected networks (See Fig. 1(a)). Here, we do not take into account the time delay of data transfer at each vertex or edge, so that all data are delivered in a unit time, regardless of the distance between any two vertices. So the load is a static variable for a given number of nodes N. Since the packets are conserved, the total load contributed by one pair is simply related to the shortest pathway length d_{ij} between them, by $\sum_k \ell_k^{(i\to j)} = d_{ij}+1$. Thus we have the sum rule for ℓ_k:

$$\sum_k \ell_k = \sum_{i\neq j}(d_{ij}+1) = N(N-1)(D+1) \sim N^2 D \ , \tag{4}$$

where D is the average of d_{ij} over all pairs, called the diameter.

The quantity, load is closely related to the one used in sociology called "betweenness centrality" (BC) to quantify how much power is centralized to people in social networks [23,24]. This is defined as follows. Let us consider a communication between two persons (i, j), represented by a pair of vertices (i, j). The communication is supposed to travel along the shortest pathway between the two. When there are more than one geodesic between them, one of them is taken with equal probability. The BC for a certain vertex is defined as the accumulated fraction of the total number of the shortest pathways passing on that vertex over all pairs. That is, the betweenness centrality at a vertex k is

$$b_k = \sum_{i \neq j} b_k^{(i,j)} = \sum_{i \neq j} \frac{C_k(i,j)}{C(i,j)}, \tag{5}$$

where $C(i, j)$ is the total number of geodesics connecting the vertices i and j, and $C_k(i, j)$ the number of those passing through the vertex k among them. The definition of BC is also illustrated in Fig. 1(b). Slightly different in the definitions, the two quantities, load or BC, behave closely and their distributions are indistinguishable within our numerical resolution. Hence we shall not distinguish them throughout this paper, unless otherwise noted explicitly.

2.1 The load distribution

Once a SF network is generated artificially or adopted from the real world, we select an ordered pair of vertices (i, j) on the network, and identify the shortest pathway(s) between them and measure the load on each vertex along the shortest pathway using the modified version of the breath-first search algorithm introduced by Newman [24] and independently by Brandes [25]. We have measured load ℓ_i of each vertex i for SF networks with various γ. It is found numerically that the load distribution $P_L(\ell)$ follows the formula,

$$P_L(\ell) \sim \ell^{-\delta}. \tag{6}$$

When the index of the vertices are ordered according to the rank of the load, we have $\ell_1 \geq \ldots \geq \ell_N$. Then, the power-law behavior of the load distribution implies that

$$\frac{\ell_i}{\sum_j \ell_j} \sim \frac{1}{N^{1-\beta} i^{\beta}}. \tag{7}$$

with

$$\delta = 1 + 1/\beta. \tag{8}$$

The relation, Eq. (7), is valid in the region, $\ell_{\min} < \ell < \ell_{\max}$, where

$$\ell_{\min} \sim \ell_{\max}/N^{\beta} \sim \begin{cases} ND & \text{if } \beta < 1 \\ ND/\ln N & \text{if } \beta = 1 \\ N^{2-\beta} D & \text{if } \beta > 1. \end{cases} \tag{9}$$

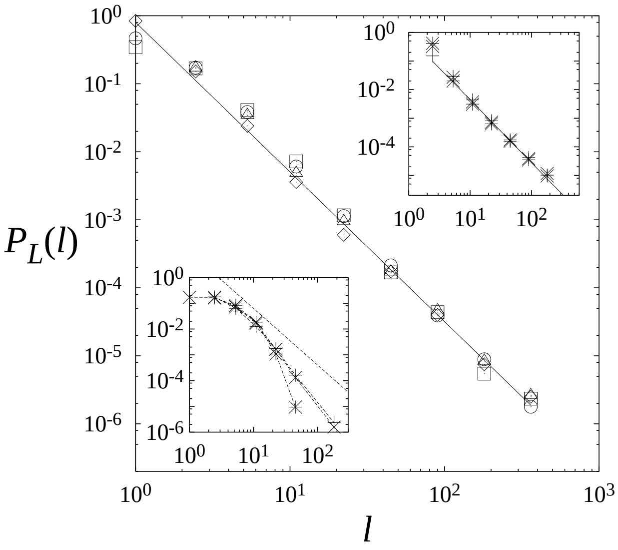

Fig. 2. Plot of the load distribution $P_L(\ell)$ versus ℓ for various $\gamma = 2.25$ (◇), 2.5 (△), 2.75 (○) and 3.0 (□) in double logarithmic scales. The data are obtained from the static model. The linear fit (solid line) has a slope -2.2. Simulations are performed for $N = 10^4$ and $m = 2$ and all data points are averaged over 10 configurations. Lower Inset: Same plot for $\gamma = 4$ (+), 5 (×), and ∞ (∗). The line having a slope -2.2 is drawn to compare the data with the case for $2 < \gamma \leq 3$. Upper Inset: Plot of $P_L(\ell)$ versus ℓ for different $m = 2$, 4 and 6, but for the same $\gamma = 2.5$.

Based on numerical measurements of load exponents for a variety of SF networks, we find that load exponent is likely to be robust, independent of the details of network structure such as the degree exponent γ in the range, $2 < \gamma < 3$, and mean degree, directionality of edge, etc. We find that SF networks are likely to be classified into two classes, say, class I and II. For the class I, the load exponent is $\delta \approx 2.2(1)$ and for the class II, it is $\delta \approx 2.0(1)$. However, see Section 2.4 for further discussions on the possibility of slight variation of δ on γ. The load exponent for the class II seems to be exactly $\delta = 2$ since it can be derived analytically for simple models. We may show that such different universal behaviors in the load distribution originate from different generic topological features of networks.

For $\gamma > 3$, however, δ depends on γ in a way that it increases as γ increases. Eventually, the load distribution decays exponentially for $\gamma = \infty$ as shown in the lower inset of Fig. 2. Thus, the transport properties of the SF networks with $\gamma > 3$ are fundamentally different from those with $2 < \gamma < 3$. Note that Eqs. (1) and (6) combined gives a scaling relation between load and degree of a certain vertex as

$$\ell \sim k^{(\gamma-1)/(\delta-1)}, \tag{10}$$

provided the ranks of vertices in load and degree do not vary significantly. See Fig. 9 and discussions in Section 3.

2.2 Real world and artificial networks investigated

The networks that we find to belong to the class I with $\delta = 2.2(1)$ include:

- (i) The co-authorship network in the field of the neuroscience, published in the period 1991-1998 [17], where vertices represent scientists and they are connected if they wrote a paper together.
- (ii) The protein interaction network (PIN) of the yeast *Saccharomyces cerevisiae* compiled by Jeong *et al.* (PIN1) [26], where vertices represent proteins and the two proteins are connected if they interact.
- (iii) The core PIN of the yeast *S. cerevisiae* by Ito *et al.* (PIN2) [27].
- (iv) The metabolic networks for 5 species of eukaryotes and 32 species of bacteria in Ref. [18], where vertices represent substrates and they are connected if a reaction occurs between two substrates via enzymes. The reaction normally occurs in one direction, so that the network is directed.
- (v) The Barabási-Albert (BA) model [19] when the number of incident edges of an incoming vertex $m \geq 2$ where $\gamma = 3$.
- (vi) The geometric growth model by Huberman and Adamic [28].
- (vii) The copying model [29] where $\gamma = 3$.
- (viii) The undirected or the directed static model [22] where $\gamma = 3$ or $\gamma_{\rm in} = \gamma_{\rm out} = 3$.
- (ix) The accelerated growth model proposed by Dorogovtsev *et al.* [30].
- (x) The stochastic model for the PIN by Solé *et al.* [31].

For both (ii) and (x), the degree distribution is likely to follow a power-law but there needs an exponential cutoff to describe its tail behavior for finite system. However, it converges to a clean power-law for (x) as system size increases, but the converging rate is rather slow [32]. Despite this abnormal behavior in the degree distribution for finite system, the load distribution follows a pure power law with the exponent $\delta \approx 2.2(1)$ in (ii) and (x). The representative load distributions for real world networks (i), (iii), and (iv) are shown in Fig. 3(a).

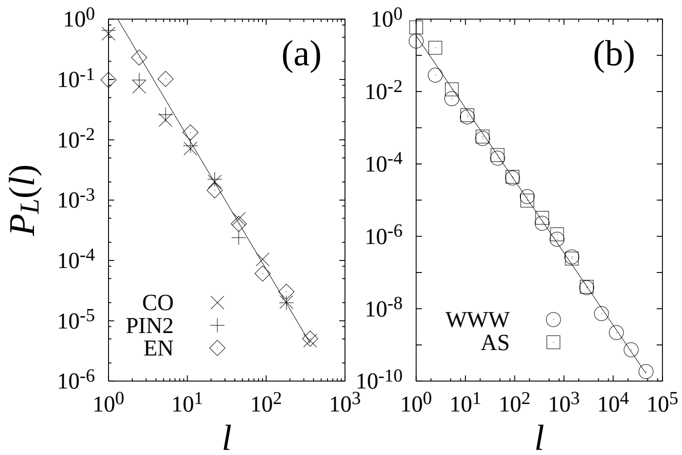

Fig. 3. The load distribution for (i), (iii), (iv), (xi), and (xiii).

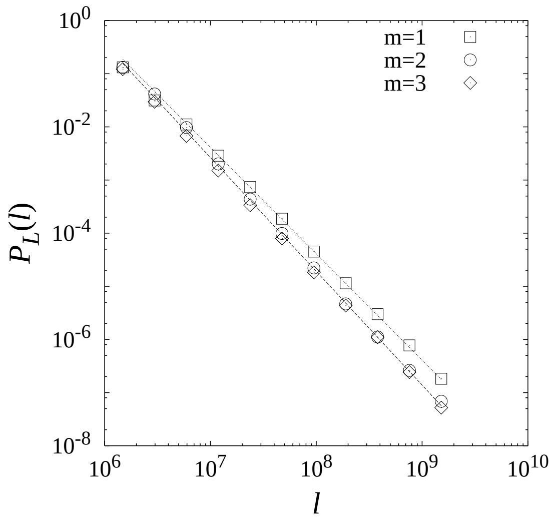

Fig. 4. The load distribution for the BA model for $\gamma = 3$ with $m = 1, 2$, and 3.

The networks that we find to belong to the class II with $\delta = 2.0$ include:

(xi) The Internet at the autonomous systems (ASes) level [33].
(xii) The metabolic networks for 6 species of archaea in Ref. [18].
(xiii) The WWW within www.nd.edu domain [9].
(xiv) The BA model with $m = 1$ for $\gamma = 3$ [19].
(xv) The deterministic model by Jung *et al.* [34].

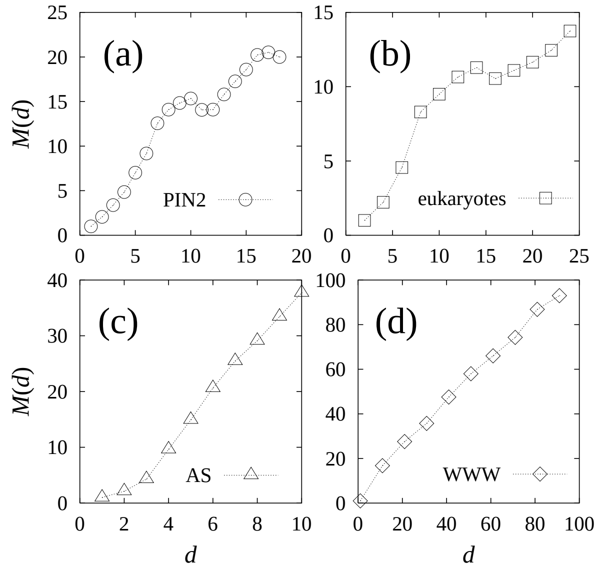

Fig. 5. The mass-distance relation for prototypical SF networks

In particular, the networks (xiv) and (xv) are of tree structure, where the edge load distribution can be solved analytically. The load distributions for real world networks (xi) and (xiii) are shown in Fig. 3(b). Since the load exponents of each class are very close numerically, one may wonder if there exist really two different universal classes apart from error bar. To make this point clear, we plot the load distributions for the BA model with $m = 1, 2$ and 3 in Fig. 4, obtained from large system size, $N = 3 \times 10^5$. We can see clearly different behaviors between the two load distributions for the cases of $m = 1$ (class II) and of $m = 2$ and 3 (class I) in the BA model with $\gamma = 3$.

2.3 Topology of the shortest pathways

To understand the generic topological features of the networks in each class, we particularly focus on the topology of the shortest pathways between two vertices separated by a distance d. Along the shortest pathways, we count the total number of vertices $\mathcal{M}(d)$ lying on these roads, averaged over all pairs of vertices separated by the same distance d. Adopting from the fractal theory, $\mathcal{M}(d)$ is called the "mass-distance" relation. We find that it behaves in different ways for each class; For the class I, $\mathcal{M}(d)$ behaves nonlinearly (Figs. 5a-b), while for the class II, it is roughly linear (Fig. 5c-d). For the networks belonging to the class I such as the PIN2 (iii) and the metabolic network for eukaryotes (iv), $\mathcal{M}(d)$ exhibits a non-monotonic behavior (Fig. 5a-b), *viz.*,

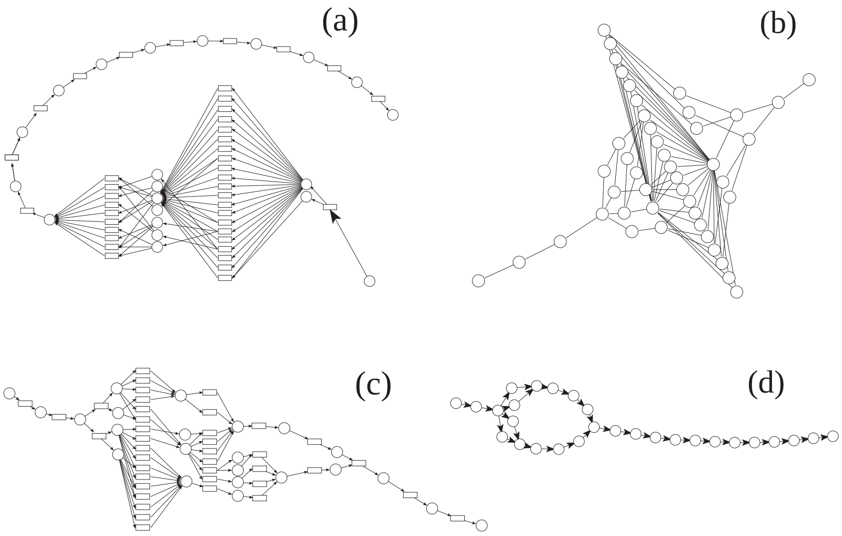

Fig. 6. Topology of the shortest pathways. (a) The metabolic network of *E. nidulans* (eukaryote) of length 26. (b) The Internet AS of length 10. (c) The metabolic network of *Methanococcus jannaschii* (Archaea) of length 20. (d) WWW of www.nd.edu of length 20. In a and c, circles denote substrates, and rectangles denote intermediate states.

it exhibits a hump at $d_h \approx 10$ for (iii) or $d_h \approx 14$ for (iv). To understand why such a hump arises, we visualize the topology of the shortest pathways between a pair of vertices, taken from the metabolic network of a eukaryote organism, *Emericella nidulans* (*EN*), as a prototypical example for the class I. Fig. 6(a) shows such a graph with linear size 26 edges ($d = 26$), where an edge between a substrate and an enzyme is taken as the unit of length. From Fig. 6(a), one can see that there exists a blob structure inside which vertices are multiply connected, while vertices outside are singly connected. What is characteristic for the class I is that the blob is localized in a small region.

For the class II, the mass depends on distance linearly, $\mathcal{M}(d) \sim Ad$ for large d (Fig. 6(c)-(d)). Despite the linear dependence, the shortest pathway topology for the case of $A > 1$ is more complicated than that of the simple tree structure where $A = 1$. Therefore, the SF networks in the class II are subdivided into two types, called the class IIa and IIb, respectively. For the class IIa, $A > 1$ and the topology of the shortest pathways includes multiply connected vertices (Figs. 6(b) and (c)), while for the class IIb, $A \sim 1$ and the shortest pathway is almost singly connected (Fig. 6(d)). Examples in real world networks in the class IIa are the Internet at the AS level ($A \sim 4.5$) and the metabolic network for archaea ($A \sim 2.0$), while that in the class IIb is the WWW ($A \sim 1.0$).

Table 1. Comparison of the load exponent for the static model with $m = 2$: δ_1 and δ_2 are the measured values using the definitions of [22] and [38], respectively. "log" and "cum" mean the measurement by log-binned and cumulative ways, respectively.

N	γ	δ_1(log)	δ_1(cum)	δ_2(log)	δ_2(cum)
1×10^4	3.0	2.30(1)	2.34(5)	2.27(3)	2.30(3)
1×10^4	2.8	2.23(3)	2.22(5)	2.17(1)	2.19(1)
1×10^4	2.5	2.15(4)	2.13(3)	2.07(1)	2.10(1)
1×10^4	2.2	2.05(5)	2.00(4)	1.96(2)	2.01(2)
1×10^5	2.5	2.13(3)	2.10(2)	2.08(1)	2.10(1)
5×10^5	2.5	2.11(2)	2.08(2)	2.08(1)	2.07(1)

The WWW is an example belonging to the class IIb. For this network, the mass-distance relation exhibits $\mathcal{M}(d) \sim 1.0d$, suggesting that the topology of the shortest pathway is almost singly connected, which is confirmed in Fig. 6(d).

2.4 Crossover in load distributions between tree and loop scale-free networks

The load exponents for various SF trees have been obtained analytically to be $\delta = 2.0$, independent of the degree exponent γ [35,36]. We investigate how the value of the exponent $\delta = 2.0$ changes as the number of loops increases. To this end, we modify the BA model in such a way that a new vertex attaches one or two edges to existing network $1 - p$ or p, respectively [37]. Then the mean number of edges emanating from a new vertex is given by $\langle m \rangle = 1 + p$. When $p = 0$, the network is a tree, and the load exponent is confirmed to be $\delta = 2.0$. In addition, we find that the load exponent increases to $\delta \simeq 2.2$ by increasing $\langle m \rangle$ to $\langle m \rangle \simeq 1.1$ at which the edges connecting different branches of the tree structure form sparse loops in a nontrivial manner. As shown in Fig. 7, the value $\delta \simeq 2.2$ at $\langle m \rangle \simeq 1.1$ turns out to be robust, largely independent of the degree exponent γ for $2 < \gamma < 3$. Such a universal behavior persists as long as $\langle m \rangle$ is smaller than a γ-dependent characteristic value, $\langle m \rangle_c$, beyond which δ depends on γ. Moreover, we find that the plateau region of $\delta \simeq 2.2$ is extended as the system size N increases as shown in Fig. 7. These data suggest that the universal behavior of δ may hold in some finite region of parameter space in the thermodynamic limit, for the sparse graph of BA model. Note that if one measures the load exponent δ at $\langle m \rangle = 2$ in finite size systems, the universal features could not be observed as Barthélemy noticed recently [38] in the static model.

Barthélemy performed extensive numerical simulations of the load distribution of the static model and the BA model, finding that the robustness of the load exponent fails even for $2 < \gamma < 3$. Note that the definition of load he used is slightly different from ours. He ignored the contributions of $N - 1$

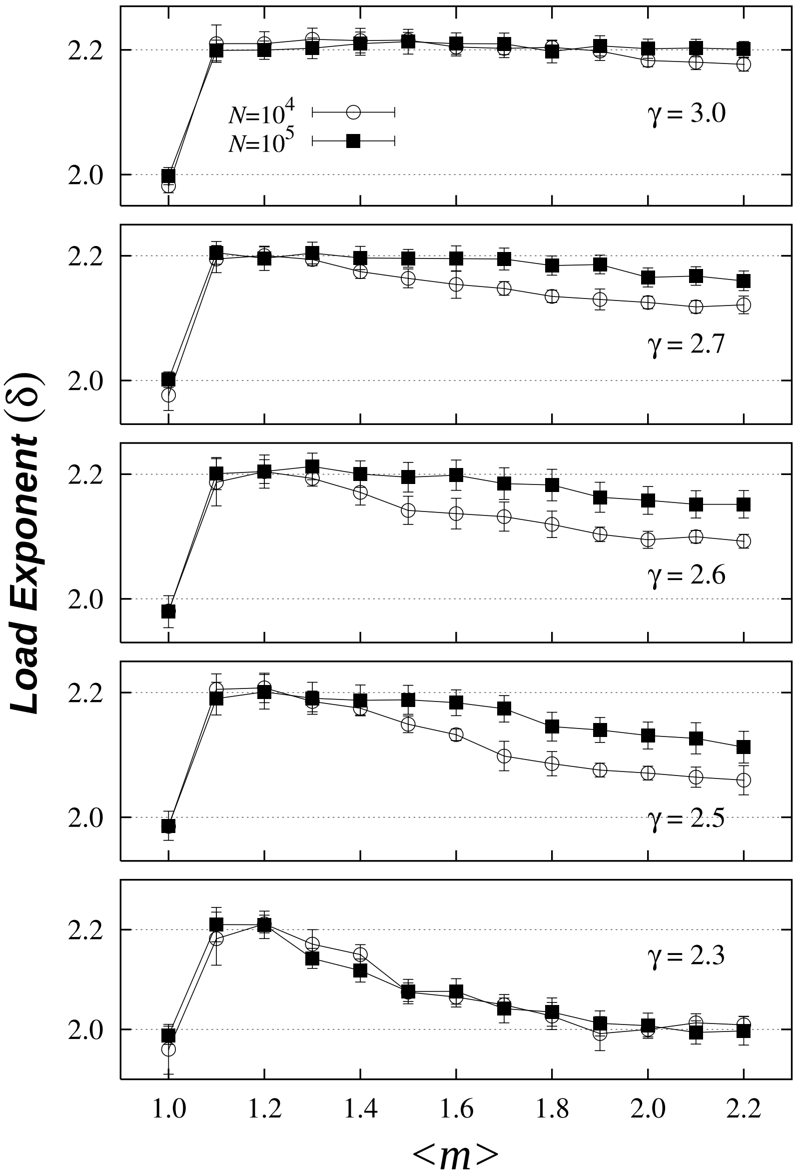

Fig. 7. The load exponent as a function of the mean number of edges $\langle m \rangle$ emanating from a new vertex for various degree exponents γ in the BA model and different system sizes, $N = 10^4$ (◯) and $N = 10^5$ (■).

packets leaving and another $N - 1$ packets arriving at the vertex to load of a certain vertex, while we counted them. The difference can be neglected for vertices with large load ℓ in the limit of $N \to \infty$, however, in finite-size systems particularly those compatible with most real world networks comprising $N = 10^3 \sim 10^4$ vertices, this difference produces somewhat different value of δ. We also performed extensive numerical simulations on a larger scale $N = 5 \times 10^5$ than the size $N = 10^4$ shown in Fig. 2 for the static model with $\gamma \approx 2.5$ and $m = 2$ following both the original definition and the new one of Ref. [38], confirming that indeed, δ turns out to have weak γ dependency. Numerical values of δ measured newly are listed in Table 1. However, the behavior of δ as a function of γ with $m = 2$ for the static model is similar to the one shown in Fig. 7 for the BA model. Therefore, whether the universal behavior of the load exponent indeed breaks down or it still holds in some range of the model parameter space is still an open problem. Finally, notice that real-world networks contain nontrivial degree-degree correlations, while the static model does not have it. Thus it would be interesting to investigate how the load exponent depends on the degree-degree correlation, which is in progress. Conclusively, the problem of the universal behavior of the load exponent is not established yet, and related analytic work is needed to clarify this issue.

3 Load-load correlation

Recently, SF networks are grouped into three types according to the behavior of the degree-degree correlation coefficient [39,40], in a different aspect from the classification of the load exponent. They are the ones exhibiting the assortative, dissortative, and neutral mixing on their degree. For the network of the assortative (dissortative) mixing, called the assortative (dissortative) network, a vertex with large degree tends to connect to vertices with large (small) degree, while for the network of the neutral mixing, there is no such tendency. The assortative network can be found in social networks such as the coauthorship network, the actor network and so on, and the dissortative network in information networks such as the Internet and the www, and in biological networks such as protein interaction networks and neural networks. While such assortative and dissortative networks appear in real world, the neutral network, *i.e.*, the network of the neutral mixing on their degree, appears in *in silico* networks such as the BA model and the copying model [29] with $\gamma = 3$.

The degree-degree correlation [41,42] was investigated in terms of the correlation function between the remaining degrees of the two vertices on each side of an edge. Here the remaining degree means the degree of that vertex minus one [39]. First one defines the joint probability $e_D(j,k)$ that the two vertices on each side of a randomly chosen link have j and k remaining degrees, respectively. Then the normalized correlation coefficient is defined as

$$r_D = \frac{1}{\sigma_D(q)^2} \sum_{j,k} jk\{e_D(j,k) - q_D(j)q_D(k)\}, \tag{11}$$

Table 2. Size N, mean degree $\langle k \rangle$, degree assortativity coefficient r_D, load assortativity coefficient r_L for a number of social networks.

Type	Name	N	$\langle k \rangle$	r_D	r_L	Ref.
Actor	Videomovie	29824	33.7	0.22	0.024	[43]
	TVminiseries	33980	73.0	0.38	0.033	[43]
	TVcablemovies	117655	55.5	0.14	0.035	[43]
	TVseries	79663	118.4	0.53	0.013	[43]
Coauthor	Neuroscience	205202	11.8	0.60	0.057	[17]
	Mathematics	78835	5.50	0.59	0.091	[17]
	cond-mat	16264	5.85	0.18	0.086	[16]
	arXiv.org	52909	9.27	0.36	0.057	[16]

where $q_D(k)$ is the normalized distribution of the remaining degree $q_D(k) = (k+1)P_D(k+1)/\sum_j jP_D(j)$, and $\sigma_D(q)^2 = \sum_k k^2 q_D(k) - [\sum_k kq_D(k)]^2$. Recently Newman called this quantity the degree assortativity coefficient [40]. For the assortative (dissortative) networks, r_D is positive (negative), and for the neutral networks, $r_D = 0$. On the other hand, the degree-degree correlation was also investigated in terms of the mean degree of neighbors of a vertex with degree k, denoted by $\langle k_{\rm nn}\rangle(k)$ [13]. For the assortative (dissortative) networks, $\langle k_{\rm nn}\rangle(k)$ increases (decreases) with increasing k, while the neutral networks, $\langle k_{\rm nn}\rangle(k)$ is independent of k.

For the BA-type model, it was shown that the load is related to the degree via the relation, Eq. (10). Thus the vertices with larger degree are much more influential to others in communications. Due to this relation, one may think that the load-load correlation would behave similarly to the degree-degree correlation. Here we show that while for the dissortative and neutral network, the load-load correlation coefficients behave similarly to the degree-degree correlation coefficients, for the assortative network, the relation Eq. (10) is nontrivial, leading to that the load-load correlation is very weakly assortative, i.e., the mean load of neighbors of a certain vertex with load ℓ_i is almost independent of ℓ_i.

To study the load-load correlation, we introduce the load-load correlation coefficient, called the load assortativity coefficient, in analogy with Eq. (11):

$$r_L = \frac{1}{\sigma_L(q)^2} \sum_{\ell,m} \ell m \{e_L(\ell,m) - P_L(\ell)P_L(m)\}, \tag{12}$$

where $e_L(\ell,m)$ is the joint probability that the loads of the two vertices of a link are ℓ and m and $\sigma_L(q)^2 = \sum_\ell \ell^2 P_L(\ell) - [\sum_\ell \ell P_L(\ell)]^2$. Moreover, similarly to $\langle k_{\rm nn}\rangle$, we define the mean load of neighbors of a vertex with load ℓ, denoted by $\langle \ell_{\rm nn}\rangle(\ell)$, through which we can check if the load-load correlation is assortative or dissortative.

We first check the load-load correlation for the network of the Internet on the level of autonomous systems as of January 2000 [33] and the so-called

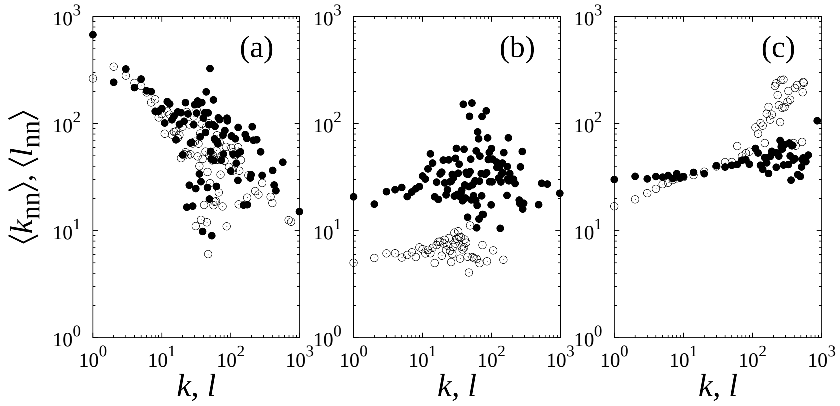

Fig. 8. Plot of $\langle k_{\rm nn}\rangle(k)$ ($\circ$) and $\langle \ell_{\rm nn}\rangle(\ell)$ ($\bullet$) for (a) the Internet on the level of autonomous systems (dissortative), (b) the non-degenerate configuration model with $\gamma = 3$ (neutral), and (c) the coauthorship network in the field of neuroscience (assortative). All data are obtained from a single configuration.

non-degenerate configuration model with $\gamma = 3$ [42,44,45], which belong to the dissortative and the neutral network, respectively. For these networks, r_L is -0.16 (< 0) and 0.02, respectively, which is close to their r_D values of -0.18 and 0.01, respectively. Moreover, $\langle \ell_{\rm nn}\rangle(\ell)$ behaves similarly to $\langle k_{\rm nn}\rangle(k)$ as shown in Fig. 8 (a) and (b). However for the assortative networks, the coauthorship network for example, r_L is considerably smaller than r_D often by one order of magnitude and is close to zero. The comparison of r_L and r_D for various social networks are tabulated in Table 2. The mean load $\langle \ell_{\rm nn}\rangle(\ell)$ of neighbors of a vertex with load ℓ increases with increasing ℓ, however, the increasing rate is very low compared with that of $\langle k_{\rm nn}\rangle(k)$, *i.e.*, it depends on ℓ very weakly (Fig. 8(c)). Such a behavior appears in other social networks too. Since load is regarded as a good measure of centrality, it implies that the mean influence of neighbors of a person is almost the same regardless of the influence of the centered person. So a person is surrounded by almost the same influential people on average no matter how influential the centered person is, although a person who is acquainted with many people is likely to connect to people who also is acquainted with many others.

To understand the abnormal behavior of the load-load correlation in detail, we examine the degree-load relation. In Fig. 9, we compare the degree-load relation $\ell(k)$ for the three types. While the relation of Eq. (10) holds for the dissortative and the neutral networks, it breaks down for large k for the assortative networks. Rather the loads of large k vertices cover wide range of values. Since the vertices with large degree are located next to each other in the assortative network, the shortest pathways between a certain pair of vertices do not necessarily pass through such nearby hubs at the same time.

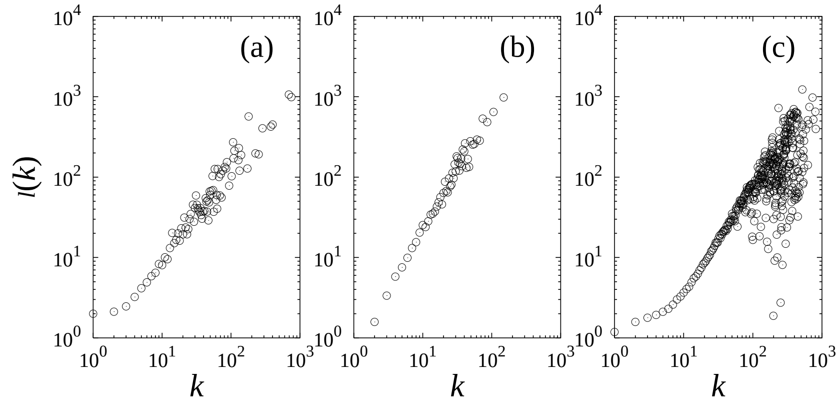

Fig. 9. Plot of the degree-load relation for (a) the Internet on the level of autonomous systems (dissortative), (b) the non-degenerate configuration model with $\gamma = 3$ (neutral), and (c) the coauthorship network in the field of neuroscience (assortative).

Thus the loads of the vertices with large k fluctuate and the degree-load correlation is nontrivial.

4 Diameter change distribution

One of the powerful ways of examining the intrinsic nature of a complex system is to observe how emerging patterns change by 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* how much the change would be definitely. Recently, Parisi has argued [46] 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 case of proteins, subject to small external perturbations such as the change in pH 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 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 [47]. The number of spins bursting 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 spins burst in this case can only be probabilistic. All these examples aptly illustrate how the concept of probabilistic prediction may apply as a new paradigm in modern science.

In this section 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 the 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 few 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,

$$P_{\rm c}(\Delta) \sim \Delta^{-\zeta} \tag{13}$$

for large Δ. Here 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 $P_{\rm c}(\Delta)$ 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 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 [48,49] 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 the computation time of order $\mathcal{O}(N^3)$.

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. 10). However, it exhibits a fat tail for $\Delta > 0$ (the inset of Fig. 10). For the static model with $\gamma = 3$, for example, the case of small diameter changes in the range $|\Delta| < 2\times10^{-4}$ occurs

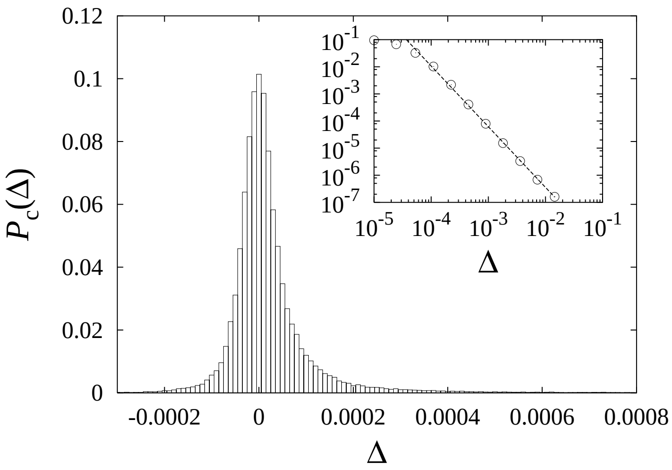

Fig. 10. Normalized histogram of the diameter changes for the static model with $\gamma = 3$ and $N = 10^4$, averaged over 10 configurations. Horizontal range is truncated for clearance, but runs up to 2×10^{-2}. Inset: Plot of $P_c(\Delta)$ in log-log scale for $\Delta > 0$. Dashed line is a fit line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

Table 3. Summary of the results for various SF networks. Tabulated for each network are the degree exponent γ, the diameter change exponent ζ, and the betweenness centrality exponent η [35].

	System	γ	ζ	η	ref.
(i)	Static model	2.2–3.0	2.2(1)	2.2(1)	[22]
(ii)	Barabási-Albert model	2.2–3.0	2.2(1)	2.2(1)	[19]
(iii)	Copying model	2.2–3.0	2.2(1)	2.2(1)	[29]
(iv)	Fitness model	2.25	2.2(1)	2.2(1)	[50]
(v)	Accelerated-growth model	3.0(1)	2.2(1)	2.2(1)	[30]
(vi)	Huberman-Adamic model	3.0(1)	2.2(1)	2.2(1)	[28]
(vii)	PIN model	–	2.2(1)	2.2(1)	[31]
(viii)	PIN of the yeast	3.2(2)	2.3(1)	2.3(1)	[51]
(ix)	Internet at the autonomous systems level	2.1(1)	1.7(1)	2.0(1)	[33]
(x)	Adaptation model	2.1	1.7(1)	2.0(1)	[14]

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 random graph and $D_0 \sim \ln N / \ln \ln N$ [52] for the BA model [19] with $\gamma = 3$. When a vertex is removed, the diameter may be reduced as $D \sim$

$\ln(N-1)$ or $D \sim \ln(N-1)/\ln\ln(N-1)$, 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. 10. 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. (13). We find that such large diameter changes are mainly due to the removal of a vertex with large degree.

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. Similar behaviors are found in other model networks (ii)–(vii) listed in Table 3. These include the SF networks showing nontrivial degree-degree correlations [39]. 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 Erdős-Rényi random networks decays exponentially.

To see such universal behavior of ζ in real world, we consider a couple of real-world networks, the PIN and the Internet. For the PIN of the yeast *S. cerevisiae* [51], we also find a power law in the diameter change distribution with an exponent $\zeta \approx 2.3(1)$ (Fig. 11), consistent with the one obtained for various model networks, including the one proposed as its own *in silico* model (vii) [31]. For the Internet at the autonomous systems level [33], the diameter change distribution again follows a power law, however, with a different exponent $\zeta \approx 1.7(1)$ (Fig. 12). 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 3]. To confirm the novel value of ζ for the Internet, we perform the same calculations for its *in silico* model, called the adaptation model [14], 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.

Interestingly, the networks (i)–(viii) in Table 3 having the diameter change exponent $\zeta \approx 2.2$ belong to the class I, and the values of ζ and δ coincide with each other within our numerical resolutions, while they are different for the class II. Empirically, the rank of a vertex in ℓ and that in Δ are likely to be the same for vertices with large degrees. If then, the relation $P_L(\ell)d\ell \sim P_c(\Delta)d\Delta$ would hold asymptotically, leading to

$$\Delta(\ell) \sim \ell^{(\delta-1)/(\zeta-1)}, \tag{14}$$

for large ℓ. Indeed, the slopes in the double logarithmic scale in the upper insets of Figs. 11 and 12 are 1.1(1) for the PIN and 1.4(1) for the Internet, respectively, consistent with the predictions from the formula, Eq. (14). Thus the two classes, the classes I and II, are also categorized by the diameter change distribution and the distinction between them can be observed more clearly through it.

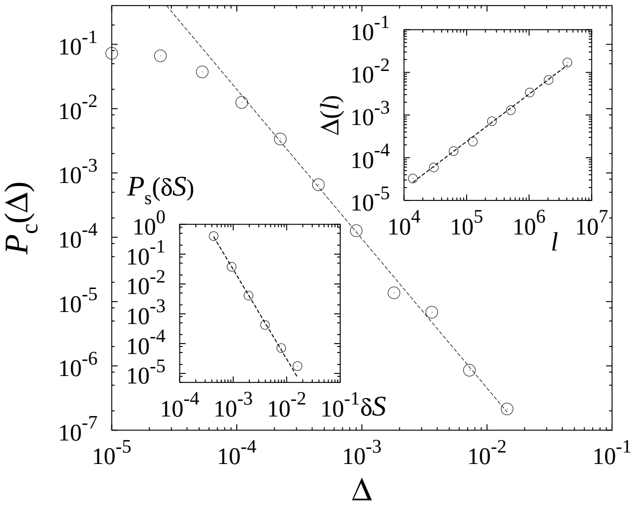

Fig. 11. The diameter change distribution $P_c(\Delta)$ for the PIN of the yeast *S. cerevisiae*. The slope of the fit line (dashed) is -2.3, drawn for the eye. Upper inset: Plot of $\Delta(\ell)$ *vs.* ℓ. The slope of the straight line is 1.1, drawn for the eye. Lower inset: The largest-cluster-size 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.

Our finding that the diameter change distribution is also classified into the classes I and II following those for load distribution may be rooted from the fact that both quantities, diameter and load, depend on universal features of the shortest pathways topology between a vertex pair in networks. When the sum rule [53], $\sum_k \ell_k \sim D$, is applied, one can see immediately that the diameter change distribution is the same as the total load change distribution. On the other hand, the networks belonging to the class II are more sparse and ramified than those in the 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 of Figs. 11 and 12, 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 δ in the class II. However, it is not clear how the power-law behavior in $P_c(\Delta)$ arises and what determines its exponent.

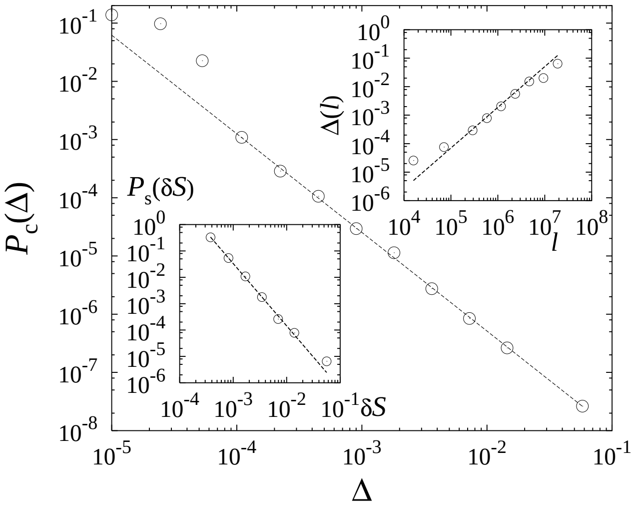

Fig. 12. The diameter change distribution $P_c(\Delta)$ for the Internet at the autonomous system level. The slope of the fit line (dashed) is -1.7, drawn for the eye. Upper inset: Plot of $\Delta(\ell)$ *vs.* ℓ. The slope of the straight line is 1.4, drawn for the eye. Lower inset: The largest-cluster-size 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.

5 Conclusions and discussion

We have introduced a quantity called *load* to study a transport property of data packet on SF network, finding that the load distribution follows a power law with exponent δ. Interestingly, the load exponent δ seems to be robust, independent of the details of network structures. So far, we have found that there exist two different classes where $\delta \approx 2.2(1)$ or 2.0. The classification is mainly rooted from generic topological features of the shortest pathways between a pair of vertices. For the former, most shortest pathways are made of multiple pathways and have a blob structure, while for the latter, it is effectively a tree. Whether the load exponent δ is strictly universal in the sense of critical phenomena and renormalization group is now an open question. Initially, we conjectured it to be so based on real world network data and topological difference of the shortest pathways. Recently, however, it was observed [38] that the universality does not hold for the static model as γ approaches 2. On the other hand, we have studied the crossover behavior of the load exponent from a tree to looped structure as the number of loops increases in the BA model. The numerical result indicates two aspects: On one hand the universal behavior may still hold in the thermodynamic limit of some *in silico* model networks, however, on the other hand the universality of the load distribution may not hold for real world class I networks. Further studies are needed to resolve this controversial issue. The contents of this review article are mainly based on our works published in [22,35,37,53–55]

This work is supported by the KOSEF Grant No. R14-2002-059-01000-0 in the ABRL program. BK would like to thank the organizing committee for inviting him to the conference.

References

1. K. Ziemelis and L. Allen, Nature **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. S. H. Strogatz, Nature **410**, 268 (2001).
4. R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
5. S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
6. M. E. J. Newman, SIAM Review **45**, 167 (2003).
7. P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. Ser. A **5**, 17 (1960).
8. A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
9. R. Albert, H. Jeong, and A.-L. Barabási, Nature **401**, 130 (1999).
10. B. A. Huberman and L. A. Adamic, Nature **401**, 131 (1999).
11. A. Broder, *et al.*, Computer Networks **33**, 309 (2000).
12. M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
13. R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
14. K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **88**, 108701 (2002).
15. S. Redner, Eur. Phys. J. B **4**, 131 (1998).
16. M. E. J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).
17. A.-L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, Physica A **311**, 590-614 (2002).
18. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvani, and A.-L. Barabási, Nature **407**, 651 (2000).
19. A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
20. P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
21. S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
22. K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
23. L. C. Freeman, Sociometry **40,** 35 (1977).
24. M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
25. U. Brandes, J. Math. Sociol. **25**, 163 (2001).
26. H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).
27. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc. Natl. Acad. Sci. USA **98**, 4569 (2000).
28. B. A. Huberman and L. A. Adamic, e-print (cond-mat/9901071) (1999).
29. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, in *Proc. IEEE FOCS 2000* (IEEE Computer Society Press, Los Alamitos, CA, 2000), pp. 57-65.

30. S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 025101(R) (2001).
31. R. Solé, R. Pastor-Satorras, E. Smith, and T. Kepler, Adv. Complex. Syst. **5**, 43 (2002).
32. J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E **66**, 055101(R) (2002).
33. Meyer, D. (2001) *University of Oregon Route Views Archive Project* (http://archive.routeviews.org).
34. S. Jung, S. Kim, and B. Kahng, *Phys. Rev. E* **65**, 056101 (2002).
35. K.-I. Goh, E. Oh, H. Jeong, B. Kahng and D. Kim, Proc. Natl. Acad. Sci. USA **99**, 12583 (2002).
36. G. Szabó, M. Alava, and J. Kertész, Phys. Rev. E **66,** 026101 (2002).
37. K.-I. Goh, C.M. Ghim, B. Kahng, and D. Kim, Phys. Rev. Lett. **91,** 189804 (2003).
38. M. Barthélemy, Phys. Rev. Lett. **91**, 189803 (2003).
39. M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
40. M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
41. P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
42. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, e-print (cond-mat/0206467) (2002).
43. http://www.imdb.com.
44. M. Molloy and B. Reed, Random Structures and Algorithms **6**, 161 (1995); Z. Burda and A. Krzywicki, Phys. Rev. E **67**, 046118 (2003).
45. Given a degree sequence $\mathcal{D} \equiv \{k_i\}$ $(i = 1, \ldots, N)$, we generate a maximally random network whose degree sequence is $\mathcal{D}$, with a constraint that any self-loops and multiple edges are forbidden. By choosing $\mathcal{D} = \left\{ \left\lfloor (N/i)^{1/(\gamma-1)} \right\rfloor \right\}$, we get a SF network with degree exponent γ, where $\lfloor \cdot \rfloor$ is the floor function.
46. G. Parisi, Physica A **263**, 557 (1999).
47. J. P. Sethna, K. A. Dahmen, and C. R. Myers, Nature **410**, 242 (2001) and references therein.
48. R. Albert, H. Jeong, and A.-L. Barabási, Nature **406**, 378 (2000).
49. Z. Liu, Y.-C. Lai, and N. Ye, Phys. Rev. E **66,** 036112 (2002).
50. G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
51. K.-I. Goh, B. Kahng, and D. Kim, e-print (q.bio-MN/0312009).
52. R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
53. K.-I. Goh, B. Kahng, and D. Kim, Physica A **318**, 72 (2003).
54. K-I. Goh, B. Kahng and D. Kim, Phys. Rev. E **67**, 017101 (2003).
55. J.-H. Kim, K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **91**, 058701 (2003).