Fractal Boundaries of Complex Networks

Jia Shao[1], Sergey V. Buldyrev[2,1], Reuven Cohen[3],
Maksim Kitsak[1], Shlomo Havlin[4], and H. Eugene Stanley[11]

[11] *Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[2] *Department of Physics, Yeshiva University, 500 West 185th Street, New York, New York 10033, USA*
[3] *Department of Mathematics, Bar-Ilan University, 52900 Ramat-Gan, Israel*
[4] *Minerva Center and Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel*
(Dated: January 17, 2008sbckjhs.tex)

We introduce the concept of the boundary of a complex network as the set of nodes at distance larger than the mean distance from a given node in the network. We study the statistical properties of the boundary nodes of complex networks. We find that for both Erdös-Rényi and scale-free model networks, as well as for several real networks, the boundary has fractal properties. In particular, the number of boundary nodes B follows a power-law probability density function which scales as B^{-2}. The size of the clusters, which are formed by the boundary nodes after removing the non-boundary nodes, also follows a power-law probability density function which scales as s^{-3}. These clusters are fractals with a fractal dimension $d_f \approx 2$. We present analytical and numerical evidence supporting these results for a broad class of networks.

Many complex networks are "small world" due to the very small average distance d between two randomly chosen nodes. Usually $d \sim \ln N$, where N is the number of nodes [1–6]. Thus, starting from a randomly chosen node following the shortest path, one can reach any other node in a very small number of steps. This phenomenon is called "six degrees of separation" in social networks [4]. That is, for most pairs of randomly chosen people, the shortest "distance" between them is not more than six. Many random network models, such as Erdös-Rényi network (ER) [1], Watts-Strogatz network (WS) [5] and random scale-free network (SF) [3, 7, 8], as well as many real networks, have been shown to possess this small-world property.

Much attention has been devoted to the structural properties of networks within the average distance d from a given node. However, almost no attention has been given to nodes which are at distances greater than d from a given node. We define these nodes as the boundary nodes of the network. An interesting question is how many "friends of friends of friends etc..." one has at distance greater than the average distance d? What is their probability distribution and what is the structure of the boundary? The boundary nodes have an important role in several cases, such as in the spread of viruses or information in a human social network. If the virus (information) spreads from one node to all its nearest neighbors, and from them to all next nearest neighbors and further on until the average distance, how many nodes do not get the virus (information), and what is their probability distribution?

In this Letter, we find theoretically and numerically that the nodes at the boundary, which are of order N, exhibit similar fractal features for many types of networks, including ER and SF models as well as several real networks. Song *et al.* [9] found that some networks have fractal properties while others do not. Here we show that almost all model and real networks including non-fractal networks have fractal features at their boundaries.

Fig. 1 demonstrates our approach and analysis. For each node, we identify the nodes at distance ℓ from it as nodes in shell ℓ. We chose a random origin node and count the number of nodes B_ℓ at shell ℓ. We see that B_1=10, B_2=11, B_3=13, etc... We estimate the average distance $d \approx 2.9$ by averaging the distances between all pairs of nodes. After removing nodes with $\ell < d = 2.9$, the network is fragmented into 12 clusters, with sizes s_3=$\{1, 1, 2, 5, 1, 3, 1, 1, 8, 1, 2, 3\}$.

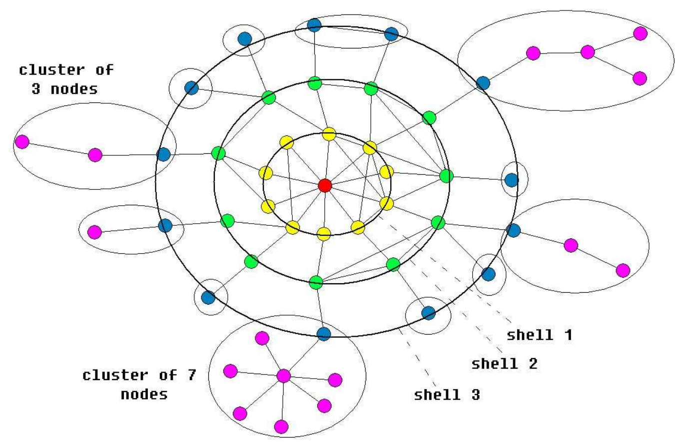

FIG. 1: (Color on line) Illustration of shells and clusters originating from a randomly chosen node, which is shown in the center (red). Its neighboring nodes are defined as shell 1, the nodes at distance ℓ are defined as shell ℓ. When removing all nodes with $\ell < 3$, the remaining network becomes fragmented into 12 clusters.

We begin our study by simulating ER and SF networks, and later present analytical proofs. Fig. 2a shows simulation results for the number of nodes B_ℓ reached from a randomly chosen origin node for an ER network. The results shown are for a single network realization of size $N = 10^6$, with average degree $\langle k \rangle = 6$ and $d \approx 7.9$ [10].

For $\ell < d$, the cumulative distribution function $P(B_l)$, which is the probability that shell ℓ has more than B_ℓ nodes, decays exponentially for $B_\ell > B^*_\ell$, where B^*_ℓ is the maximum typical size of shell ℓ [11]. However, for $\ell > d$, we observe a clear transition to a power law decay behavior, where $P(B_\ell) \sim B_\ell^{-\beta}$, with $\beta \approx 1$ and the pdf of B_ℓ is $\tilde{P}(B_\ell) \equiv dP(B_\ell)/dB_\ell \sim B_\ell^{-2}$. Thus, our results suggest a broad "scale-free" distribution for the number of nodes at distances larger than d. This power law behavior demonstrates the fractal nature of the boundaries of network, suggesting that there is no characteristic size and a broad range of sizes can appear in a shell at the boundaries. Further fractal features of the boundary structure will be shown below.

In SF networks, the degrees of the nodes, k, follow a power law distribution function $q(k) = ck^{-\lambda}$, where $c \approx (\lambda-1)k_m^{\lambda-1}$ and k_m is the minimum degree of the network, which we chose here to be 2. The largest degree K is the natural upper cutoff, $K \approx k_m N^{1/(\lambda-1)}$ [12, 13]. Fig. 2b shows, for SF networks with $\lambda = 2.5$, similar power law results, $P(B_\ell) \sim B_\ell^{-\beta}$ for $\ell > d$ as for ER, with a similar power $\beta \approx 1$. We find similar results also for $\lambda > 3$ (not shown).

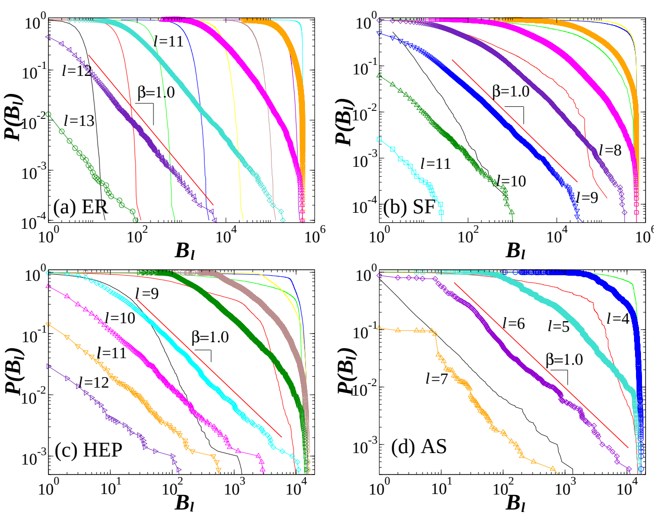

FIG. 2: The cumulative distribution function, $P(B_\ell)$, for two random network models: (a) ER network with $N = 10^6$ nodes and $\langle k \rangle = 6$, and (b) SF network with $N = 10^6$ nodes and $\lambda = 2.5$, and two real networks: (c) the High Energy Particle (HEP) physics citations network and (d) the Autonomous System (AS) Internet network. The shells with $\ell > d$ are marked with their shell number. The thin lines from left to right represent shells ℓ =1, 2 ... respectively, with $\ell < d$. For $\ell > d$, $P(B_\ell)$ follows a power-law distribution $P(B_\ell) \sim B_\ell^{-\beta}$, with $\beta \approx 1$ (corresponding to $\tilde{P}(B_\ell) \sim B_\ell^{-2}$ for the pdf). The appearance of a power law decay only happens for ℓ larger than $d \approx 7.9$ for ER and $d \approx 4.7$ for the SF network. The straight lines represent a slope of -1.

To test how general is our finding, we also study several real networks (Figs. 2c, 2d), including the High Energy Particle (HEP) physics citations network [14] and the Autonomous System (AS) Internet network [15, 16]. Our results suggest that the fractal properties of the boundaries appear also in both networks, with similar values of $\beta \approx 1$ for $\ell > d$ [17].

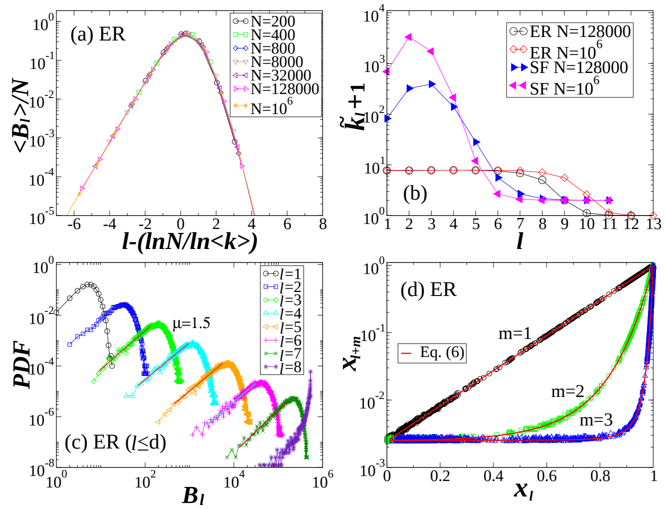

FIG. 3: (Color on line) (a) Normalized average number of nodes at shell ℓ, $\langle B_\ell \rangle/N$ as a function of $\ell - \ln N/\ln\langle k \rangle$ for ER network with $< k >$=6. For different N, the curves collapse. (b) $\tilde{k}_\ell + 1$, which is $\langle k_\ell^2 \rangle / \langle k_\ell \rangle$, as function of ℓ shown for both ER and SF network. (c) The probability distribution function $\tilde{P}(B_\ell)$ in shells $\ell \le d$ for ER network. For small values of B_ℓ, $\tilde{P}(B_\ell) \sim B_\ell^{\mu}$, where μ depends on the $\langle k \rangle$ of the network (Eq. (4)). (d) The fraction of nodes outside shell $\ell + m$, $x_{\ell+m}$, as a function of x_ℓ for ER network, where x_ℓ is calculated for any possible ℓ. The (red) lines represent the theoretical iteration function (Eq. (6)).

Next we ask how many nodes are on average at the boundaries? Are they a finite fraction of N, or less? In Fig. 3a, we study the mean number $\langle B_\ell \rangle$ in shell ℓ, and plot $\langle B_\ell \rangle/N$ as function of $\ell - \ln N/\ln\langle k \rangle$ for different values of N for ER network, where N denotes the size of the giant component of the network. The term $\ln N/\ln\langle k \rangle$ represents the average distance d of the network [2]. We find that, for different values of N, the curves collapse, supporting a relation independent of network size N. Since $\langle B_\ell \rangle/N$ is apparently constant and independent of N, it follows that $\langle B_\ell \rangle \sim N$, i.e., a finite fraction of N nodes appear at each shell including shells with $\ell > d$. We find similar behavior for SF network with $\lambda = 3.5$ (not shown). The branching factor [12] of the network is $\tilde{k} = \langle k^2 \rangle / \langle k \rangle - 1$, where the averages are calculated for the entire network. Similarly, we define $\tilde{k}_\ell = \langle k_\ell^2 \rangle / \langle k_\ell \rangle - 1$, where the averages are calculated only for nodes in shell ℓ. Above the average distance, $\tilde{k}_\ell + 1$ decreases with ℓ for both ER and SF networks (Fig. 3b). Thus, at the shells where power law behavior of $P(B_\ell)$ appears (Fig. 2), the nodes have much lower $\tilde{k}_\ell + 1$ compared with the entire network. The approach of $\tilde{k}_\ell + 1$ to 1 (ER network) and 2 (SF network) is consistent with a critical behavior at the boundaries of the network [12].

Fig. 3c shows that $\tilde{P}(B_\ell)$ for $\ell < d$ and small values of

B_ℓ increase as a power law, $\tilde{P}(B_\ell) \sim B_\ell^\mu$ for ER network, where μ depends on $\tilde{k}$ (supporting the theory developed below). We define the fraction of nodes outside shell m as $x_m = 1 - (\sum_{\ell=1}^{m} B_\ell)/N$. There exists a functional relation which is independent of ℓ between any two x_ℓ and $x_{\ell+m}$ $(m = 1, 2, 3...)$, for ER network in Fig. 3d. Figs. 3c, 3d provide empirical evidences for the theory developed below.

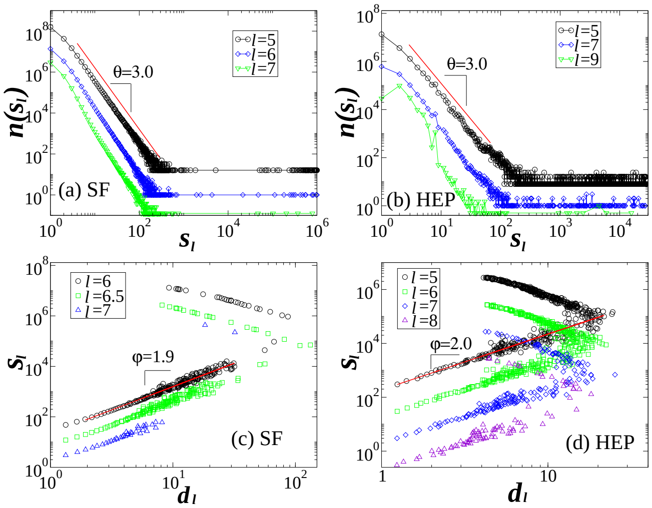

FIG. 4: The number of clusters of sizes s_ℓ, $n(s_\ell)$, as function of s_ℓ after removing nodes within shell ℓ for: (a) SF network with $N = 10^6$ and $\lambda = 2.5$, (b) HEP citations network, and s_ℓ as function of average distance d_ℓ of the clusters for (c) SF network with N=10^6 and $\lambda = 2.5$, (d) HEP citations network. The relation between $n(s_\ell)$ and s_ℓ is characterized by a power law, $n(s_\ell) \sim s_\ell^{-\theta}$, with $\theta \approx 3$. Also, s_ℓ is power law with d_ℓ, $s_\ell \sim d_\ell^\varphi$, with $\varphi \approx 2$.

Next, we study the structural properties of the boundary. Removing all nodes that are within a distance $\ell > d$ (not including shell ℓ), the network will become fragmented into several clusters (see Fig. 1). We denote the size of those clusters as s_ℓ, the number of clusters of size s_ℓ as $n(s_l)$, and the average distance in the clusters as d_ℓ [21]. We find $n(s) \sim s^{-\theta}$, with $\theta \approx 3.0$ (Figs. 4a and 4b). Similar relations are also found for ER and other real networks. The relation between the size of the clusters s_ℓ and their mean distance d_ℓ is shown in Figs. 4c and 4d, for SF ($\lambda = 2.5$) and HEP citations networks respectively. These plots suggest a power law relation, $s_\ell \sim d_\ell^\varphi$, with $\varphi \approx 2$. It indicates that the clusters at the boundaries are fractals with fractal dimension $d_f = 2$ as perculation clusters at criticality [22]. Note that, for very large clusters their average distances d_ℓ decrease with size, suggesting that the largest clusters are not fractals. We find that the fractal dimension is $d_f = \varphi \approx 2$ also for ER, SF with $\lambda = 3.5$ and some real networks.

Next we present analytical derivations supporting the above numerical results. We denote the degree distribution of a network as $q(k)$. The probability of reaching a node with k outgoing links through a link is $\tilde{q}(k) = (k+1)q(k+1)/\langle k\rangle$. We define the generating function of $q(k)$ as $G_0(x) \equiv \sum_{k=0}^{\infty} q(k)x^k$, the generating function of $\tilde{q}(k)$ as $G_1(x) = \sum_{k=0}^{\infty} \tilde{q}(k)x^k = G_0'(x)/\langle k\rangle$. For ER networks we have $G_0(x) = G_1(x) = e^{\langle k\rangle(x-1)}$. The generating function for the number of nodes, B_m, at the shell m is [23]:

$$\tilde{G}_m(x) = G_0(G_1(...(G_1(x)))) = G_0(G_1^{m-1}(x)), \qquad (1)$$

where $G_1(G_1(...)) \equiv G_1^{m-1}(x)$ is the result of applying $G_1(x)$, $m-1$ times. $\tilde{P}(B_m)$, which is the pdf of B_m, is the coefficient of x^{B_m} in the Taylor expansion of $\tilde{G}_m(x)$.

For shells with large m but still much smaller than d, we expect [23] that the number of nodes will increase by a factor of $\tilde{k}$. Hence, we conclude that $G_1^{m-1}(x)$ converges to a function of the form $f((1-x)\tilde{k}^m)$ for large m ($m << d$), and $f(x)$ satisfies the functional relation:

$$G_1(f(y)) = f(y\tilde{k}), \qquad (2)$$

where $y = 1 - x$.

The solution of $G_1(f_\infty) = f_\infty$ gives the probability that a link is not connected to the giant component of the network [24]. We can assume an asymptotic functional form, $f(y) = f_\infty + ay^{-\delta} + 0(y^\delta)$. Expanding both sides of Eq. (2) we obtain:

$$G_1(f_\infty) + G_1'(f_\infty)ay^{-\delta} = f_\infty + a\tilde{k}^{-\delta}y^{-\delta} + 0(y^\delta). \qquad (3)$$

Since $G_1(f_\infty) = f_\infty$, we have $\delta = -\ln G_1'(f_\infty)/\ln \tilde{k}$.

If $q(1) = 0$ and $q(2) \neq 0$, from $G_1(f_\infty) = f_\infty$, we have $f_\infty = 0$ and $G_1'(f_\infty) = 2q(2)/\langle k\rangle$. If $q(2) = q(1) = 0$, then $\delta = \infty$, which indicates that $f(y)$ has an exponential singularity. Therefore, networks with minimum degree $k_m \geq 3$ do not exhibit the following properties for $m << d$, and therefore have no fractal boundaries.

Applying the Tauberian theorem [25] to $f(y)$, which has a power law singularity, we conclude that the Taylor expansion coefficient of $\tilde{G}_m(x) = G_0(f((1-x)\tilde{k}^{m-1}))$, $P(B_m)$, behaves as B_m^μ with an exponential cutoff at $B_m^* \sim \tilde{k}^m$. When $q(1) \neq 0$ and $q(2) \neq 0$, we have $\mu = \delta - 1$ and when $q(1) = 0$ and $q(2) \neq 0$, we have $\mu = 2\delta - 1$. Thus the distribution of the number of nodes in the shell m with $m << d$ has a power law tail for small values of B_m:

$$P(B_m) \sim B_m^\mu. \qquad (4)$$

For ER network, Eq. (4) is supported by simulations for $m \leq d$ in Fig. 3c.

The above considerations are correct only for $m < d$, for which the depletion of nodes with large degree in the network is insignificant.

In the network, the shells behave almost deterministically and there exists a functional relation between any two shell m and shell n with $n > m$ (a detailed proof will be given elsewhere):

$$x_n = G_0(G_1^{n-m}(G_0^{-1}(x_m))), \qquad (5)$$

where $x_m = 1 - (\sum_{\ell=1}^{m} B_\ell)/N$ is the fraction of nodes outside shell m.

For ER networks, Eq. (5) yields:

$$x_{\ell+1} = e^{\langle k\rangle(x_\ell - 1)} = \Sigma_{\ell=0}^{\infty} q(k) x_\ell^k, \qquad (6)$$

which is valid for all possible ℓ. We test it in Fig. 3d.

When $m << d$ and $n >> d$, using the same considerations as before it can be shown that:

$$x_n = [a\tilde{k}(1 - x_m)]^{-\mu-1} + x_\infty, \qquad (7)$$

where $x_\infty = G_0(f_\infty) = f_\infty$, a is a constant.

Based on Eqs. (4) and (6), expressing x_m and x_n in terms of B_m and B_n, we find that for $m << d$ and $n >> d$, $B_n \sim B_m^{-\mu-1}$. Using $P(B_n)dB_n = P(B_m)dB_m$, we obtain $P(B_n) \sim B_n^{-1-\mu/(\mu+1)-1/(\mu+1)} = B_n^{-2}$, supporting the numerical findings in Fig. 2.

These results are rigorous when $\tilde{k}$ exists and when the minimum degree $k_m \leq 2$. For SF networks with $\lambda < 3$, $\tilde{k}$ diverges for $N \to \infty$. But for finite N, $\tilde{k}$ still exists. Thus the above results can also be applied to the case of $\lambda < 3$. For both ER and SF networks with $k_m \geq 3$, the power law of $P(B_n)$ with $n >> d$ cannot be observed, as we indeed confirm by simulations.

The cluster size distribution in percolation at some concentration p close to p_c is determined using the formula [12]:

$$P_p(s > S) \sim S^{-\tau+1} \exp(-S|p - p_c|^{-1/\sigma}) \,. \qquad (8)$$

In the case of random networks the percolation threshold is given by $p_c = 1/\tilde{k}$. In the exterior of the shell n ($n >> d$), we can estimate $|p - p_c| \sim (\tilde{k}(x_n) - 1)/\tilde{k}$, where $\tilde{k}(x_n)$ is calculated from nodes in the exterior of the shell n.

The cluster size distribution can be estimated by considering introducing a sharp exponential cutoff at $s = S_n^* \sim |\tilde{k}(x_n) - 1|^{-\frac{1}{\sigma}}$, so that $P_n(s > S) \sim S^{-\tau+1} P(S_n^* > S)$, where $P(S_n^* > S)$ is the probability for a given shell to have $S_n^* > S$.

Since $x_n - x_\infty$ has a smooth power law distribution and $\tilde{k}(x_\infty) < 1$, $|\tilde{k}(x_n) - 1| < S^{-\sigma} = \varepsilon$, it is proportional to ε. Thus $P(S_n^* > S) \sim S^{-\sigma}$ and $P_n(s > S) = S^{-\tau+1-\sigma}$. Therefore the cluster size distribution follows $n(s) \sim s^{-(\tau+\sigma)}$.

For ER networks and SF networks with $\lambda > 4$, $\tau = 2.5$ and $\sigma = 0.5$, the above derivations lead to $n(s) \sim s^{-3}$. For SF networks with $2 < \lambda < 4$, $\tau = (2\lambda - 3)/(\lambda - 2)$ and $\sigma = |\lambda - 3|/(\lambda - 2)$ [22]. Thus, for $\lambda > 3$, there will be $n_s \sim s^{-3}$ for SF networks. We conjecture $n_s \sim s^{-3}$ even for $2 < \lambda < 3$, although in this case $\tilde{k}(x_n)$ does not exist and the above derivations are not valid. Our numerical simulations support these results in Fig. 4a, b.

[1] P. Erdös and A. Rényi, Publ. Math. **6**, 290 (1959); Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[2] B. Bollobás, *Random Graphs* (Academic, London, 1985).

[3] R. Albert and A.-L.Barabási, Rev. Mod. Phys. **74**, 47(2002).

[4] S. Milgram, Psychol. Today **2**, 60-67 (1967).

[5] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[6] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).

[7] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: from Biological nets to the Internet and WWW* (Oxford University Press, New York, 2003).

[8] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: a statistical physics approach* (Cambridge University Press, 2004).

[9] C. Song *et al.*, Nature (London) **433**, 392 (2005); Nature Physics **2**, 275 (2006).

[10] Different realizations yield similar results. In one realization, a certain fraction of nodes are randomly taken to be origin. The histogram is obtained from B_ℓ belonging to different origin nodes.

[11] The behavior of the pdf of B_ℓ for $\ell < d$ will be discussed later and is shown in Fig. 3c.

[12] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).

[13] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 062101 (2001).

[14] Derived from the HEP section of arxiv.org; *http://vlado.fmf.uni−lj.si/pub/networks/data/hep−th/hep−th.htm* (website of Pajek).

[15] Y. Shavitt and E. Shir, DIMES - Letting the Internet Measure Itself, *http://www.arxiv.org/abs/cs.NI*/0506099

[16] S. Carmi *et al.*, PNAS **104**, 11150 (2007).

[17] We also find similar results (not shown here) for protein interaction yeast network [18, 19] and for the E.coli cellular network [18, 20].

[18] *http://www.nd.edu/ alb/* (homepage of A.L.Barabási).

[19] H. Jeong *et al.*, Nature (London) **411**, 41 (2001).

[20] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).

[21] Fractional shells with $\ell + a$ ($0 < a < 1$) are used here by removing nodes within shell ℓ and a fraction a of nodes at shell ℓ.

[22] R. Cohen *et al.*, Phys. Rev. E **66**, 036113 (2002); Chap. 4 in *Handbook of graphs and networks*, Eds. S. Bornholdt and H. G. Schuster (Wiley-VCH, 2002)

[23] M. E. J. Newman *et al.*, Phys. Rev. E **64**, 026118 (2001).

[24] L. A. Braunstein *et al.*, Int. J. Bifurcation and Chaos **17**, 2215 (2007).

[25] Weiss, G. H. *Aspects and Applications of the Random Walk* (North Holland Press, Amsterdam, 1994).