PHYSICAL REVIEW E **77**, 045101(R) (2008)

Self-affine fractals embedded in spectra of complex networks

Huijie Yang,[1,2] Chuanyang Yin,[1,3] Guimei Zhu,[1,3] and Baowen Li[1,4,*]

[1]*Department of Physics and Centre for Computational Science and Engineering, National University of Singapore, Singapore 117542*

[2]*School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China*

[3]*Department of Modern Physics, University of Science and Technology of China, Hefei Anhui 230026, China*

[4]*NUS Graduate School for Integrative Sciences and Engineering, Singapore 117597, Republic of Singapore*

(Received 29 August 2007; revised manuscript received 10 March 2008; published 18 April 2008)

The scaling properties of spectra of real world complex networks are studied by using the wavelet transform. It is found that the spectra of networks are multifractal. According to the values of the long-range correlation exponent, the Hust exponent H, the networks can be classified into three types, namely, $H>0.5$, $H=0.5$, and $H<0.5$. All real world networks considered belong to the class of $H\geq 0.5$, which may be explained by the hierarchical properties.

DOI: 10.1103/PhysRevE.77.045101 PACS number(s): 89.75.Hc, 05.45.Df

Complex networks have attracted increasing attention in recent years due to their relevance to diverse problems in physical, biological, and social sciences [1–3]. The primary purpose is to understand the relations between the underlying structures, dynamics, and functions. Generally, the dynamical processes as the transport of mass, energy, signal, and/or information occur at differen structure scales. The organization patterns at different scales may provide a reasonable solution to the problems.

Song *et al.*[4] found that the World Wide Web (WWW), social, protein-protein interaction (PPI), and cellular networks are fractal under a length-scale transform, namely, one can define a topological box in which the shortest path between each pair is less than l_B, the size of the box. The fractal behavior implies a power-law relation between the minimum number of boxes, N_B, needed to cover the entire network and the box size, $N_B(l_B)\sim l_B^{-d_B}$. d_B is the fractal dimension.

Detailed works have been done on the coverage methods [5]. It is shown that finding the minimum number of boxes to cover networks can be mapped to the graph coloring problem in the NP-complete complexity class, and the well-established algorithms in the coloring problem provide a solution close to optimal. A random burning-based algorithm is also proposed due to a number of other benefits [6].

A network with N identical nodes is described by an adjacent matrix A whose elements $A_{ij}=1$ or 0 if the nodes i and j are connected and disconnected, respectively. By mapping the nodes and the edges to atoms and bonds, the network can be regarded as a large cluster [7]. The Huckel Hamiltonian of the large cluster reads $\epsilon I+\eta A$, where ϵ and η are the site energy and the hopping integral, respectively. Generally, we can set $\epsilon=0$ and $\eta=1$, that is, the Hamiltonian is A. The spectrum of the network is defined as the rank ordered eigenvalues of A, namely, $E=\{E_1\leq E_2\leq\cdots\leq E_N\}$.

The topological structure of the network determines the spectrum. The invariance properties embedded in the spectrum in turn reflect the topological symmetries of the network. It is well known that the fractal structures of aperiodic crystals lead to the fractal behaviors of the corresponding spectra (for a detailed review, see Ref. [8] and the references therein). An interesting question is then, how the fractal structures of networks affect the corresponding spectra. In this paper, we shall detect the scaling properties embedded in spectra of networks.

The wavelet transform (WT) [9] is used to detect the scaling properties. We consider the nearest-neighbor level spacing series $L=\{L_i=E_{i+1}-E_i, i=1,2,\ldots,N-1\}$. The WT of the series L can be calculated as $T(s,a)=\frac{1}{a}\sum_{i=1}^{N-1}L_i g(\frac{i-s}{a})$. g is the wavelet and a the given scale. The wavelet transform can remove effectively polynomial trends along the series.

The series under consideration can be decomposed into many subsets characterized by different local Hurst exponents, which quantify the local singular behavior and thus relate to the local scaling of the series. Traditionally, the local Hurst exponents are evaluated through the modulus of the maximal values of $T(s,a)$ at each point in the series. We denote the positions of the WT maximum with $\{s_1,s_2,\ldots s_M\}$. In the long scale limit, the partition function is expressed as

$$Z(a,q)=\sum_{s=s_1}^{s_M}|T(s,a)|^q\sim a^{\tau(q)}. \tag{1}$$

For positive and negative q, $\tau(q)$ reflects the scalings of the large and small fluctuations, respectively.

If $\tau(q)$ is a straight line, the analyzed series contains only linear correlations (monofractal) and its slope represents the Hurst exponent. If $\tau(q)$ is a nonlinear function, the series is called multifractal, since different subsets of the series exhibit different local Hurst exponents. In order to characterize this multifractal, one considers the fractal dimensions of the subsets of the series that is characterized by $\alpha(q)$, which is related to $\tau(q)$ by a Legendre transform, $D(h)=qh-\tau(q), h=\frac{d\tau(q)}{dq}$. The width of this function for $q\rightarrow\pm\infty$ is a measure for the strength of multifractal, $\Delta\alpha=\alpha_{max}-\alpha_{min}$.

However, the numerical derivative of $\tau(q)$ in this method may induce unacceptable errors to $\Delta\alpha$. Thus, we employ a functional form fitted to $\tau(q)$ suggested by Kantelhardt *et al.* [10],

$$\tau(q)=-\ln(x^q+y^q)/\ln 2. \tag{2}$$

The distribution width of the Hurst exponent is given by

*phylibw@nus.edu.sg

$$\Delta\alpha = |\ln x - \ln y|/\ln 2. \tag{3}$$

Sometimes the bifractal is required to obtain $\Delta\alpha$. For a bifractal series $\tau(q)$ is characterized by two distinct slopes α_1 and α_2,

$$\tau(q) = \begin{cases} q\alpha_1 - 1, & q \le q_x, \\ q\alpha_2 + q_x(\alpha_1 - \alpha_2) - 1, & q > q_x, \end{cases} \tag{4}$$

or

$$\tau(q) = \begin{cases} q\alpha_1 + q_x(\alpha_2 - \alpha_1) - 1, & q \le q_x, \\ q\alpha_2 - 1, & q > q_x. \end{cases} \tag{5}$$

We can obtain the multifractal strength, $\Delta\alpha = |\alpha_1 - \alpha_2|$. These forms can be derived from a modification of the multiplicative cascade model [11].

In the multifractal case, one conventionally refers to the second moment as the Hurst exponent, i.e.,

$$H = [\tau(2) + 1]/2. \tag{6}$$

For $H > 0.5$, the levels will tend to form local clusters with small level spacings in different scales, while for $H < 0.5$ these clusters cannot be formed. The critical value of $H = 0.5$ corresponds to a series showing that the corresponding integrated series behaves similar to a random walk. These characteristics are induced obviously by the structures of networks generated by different mechanisms. Therefore, the exponent H can be used as a criterion to classify networks into three categories with $H < 0.5$, $H = 0.5$, and $H > 0.5$, respectively.

Theoretically, we should have $\tau(0) = -1$ while the calculated values may deviate slightly from it. The deviation $\Delta\tau_0 = 1 - |\tau(0)|$ can be used as the estimation of the error of $\tau(2)$. The corresponding error of H is

$$\delta H = \Delta\tau(0)/2. \tag{7}$$

In each application reported below we have used the real analytic wavelet $g^{(n)}$ among the class of derivatives of the Gaussian function. The polynomial trends up to n order can be removed. We present the results by using the parameter value $n = 4$. Calculations with higher orders ($n = 5$ and 6) lead to almost the same results.

Randomizing L, we detect also the scaling behaviors embedded in the resulting series (called shuffled series) as a comparison. The partition function $Z(a, q)$ are calculated by using the software provided in PHYSIOTOOKIT [12]. The integrated series, i.e., the spectrum E is used as the input data. The relation in Eq. (6) is also checked by using the detrended fluctuation analysis software.

We examine the scaling behaviors for the spectra of some real world networks [4]. The cellular networks consider the cellular functions as intermediate metabolism and bioenergetics, information pathways, electron transport, and transmembrane transport. The direct edges are replaced simply with nondirected edges. Generally, we find the spectra of real world networks to be multifractal. We present in Fig. 1 the result for the E. Coli cellular network. The Hurst exponent distributes in a wide range, $\Delta\alpha = 1.0$. The long-range correlation exponent is 0.5. For the actor network, we consider

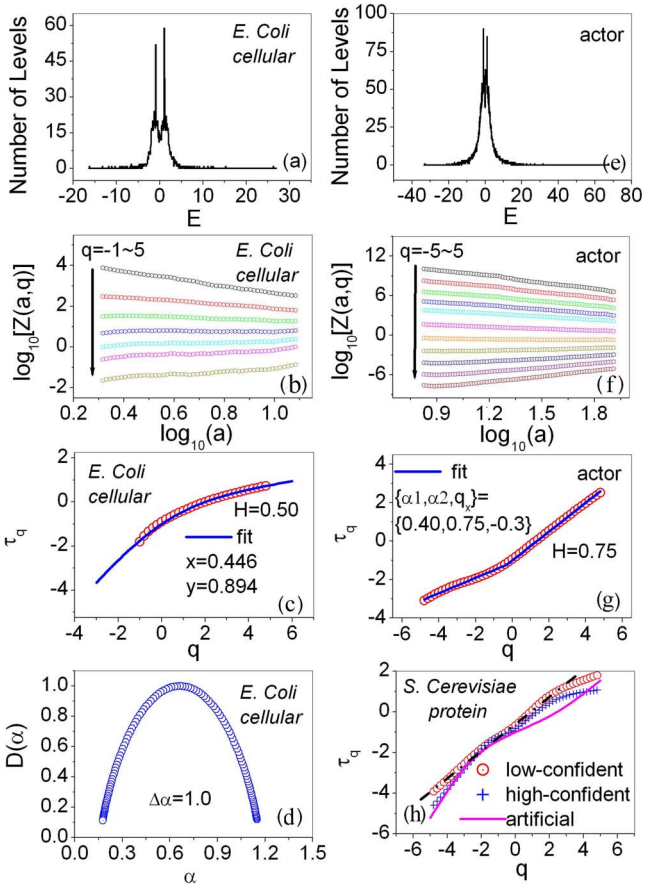

FIG. 1. (Color online) Self-affine properties embedded in spectra of real world networks. (a), (b) Histograms of the levels of the E. Coli cellular network and the actor subnetwork (containing the nodes numbered 1–8 000). (c) (d) Partition functions. (e) (f) Scaling exponent τ_q as a function of q. The E. Coli cellular network behaves multifractal. The actor subnetwork behaves bifractal, and Eq. (5) is used to obtain $\Delta\alpha$. (g) (h) The relations of τ_q versus q for the high confident, low confident, and artificial versions of the S. cerevisiae protein-protein interaction network. High confidence may not necessarily imply high quality. The result for the artificial network is an average over 20 realizations. A dashed line is added as reference, the slope of which is 0.66. The partition functions are shifted to avoid overlapping.

only the subnetwork containing the nodes numbered 1–8 000. The spectrum for this network behaves bifractally, and the long-range correlation exponent is 0.75, as given in Figs. 1(d) and 1(e).

For the S. cerevisiae protein-protein interaction network, we consider two versions of the database. One is investigated by [4], which contains 1381 nodes and 2493 edges. The other one is from [13,14], which has 1037 nodes and 1058 edges. The edges in this version are high confident. They are called low-confident and high-confident networks, respectively. As shown in Fig. 1(c), the addition of the so-called low-confident edges in the low-confident network makes $\tau(q)$ versus q significantly closer to a linear relation. The slope of the black dashed line is $H = 0.66$.

This change of the relations of τ_q versus q for the high-confident and low-confident networks may be caused simply

TABLE I. The self-affine fractals embedded in spectra of real world, WSSW, and BASF networks. For the real world networks the values of H are basically in the range of $H \geq 0.45 \approx 0.5$, while that for WSSW and BASF networks are significantly smaller, namely, $H \leq 0.3$. We present only the results for networks whose partition functions meet the scaling relation in Eq. (2) in a wide range of q, namely $\Delta q \geq 5$.

Networks		$x/y/\Delta\alpha$ or $\{\alpha_1, \alpha_2, q_x\}$	$H/\delta H$	Networks		$x/y/\Delta\alpha$ or $\{\alpha_1, \alpha_2, q_x\}$	$H/\delta H$
WWW [3]		{0.00, 0.79, 0.00}	0.97/0.08	WSSW	$p_r=0.00$	0.50/0.50/0.00	1.02/0.01
Actor [3]		{0.40, 0.75, −0.3}	0.75/0.03		$p_r=0.03$	0.81/0.81/0.00	0.31/0.01
PPI [12]	*D. melanogaster*	{0.51, 0.66, 0.40}	0.66/0.01		$p_r=0.12$	0.76/0.95/0.33	0.22/0.01
	C. elegans	0.41/0.67/0.73	0.85/0.07		$p_r=0.15$	0.66/1.00/0.61	0.24/0.01
Cellular [3]	*B. burgdorferi*	0.61/0.61/0.00	0.72/0.07		$p_r=0.21$	0.76/1.00/0.40	0.17/0.01
	A. aeolicus	0.42/0.80/0.94	0.64/0.01		$p_r=0.24$	0.70/1.00/0.50	0.21/0.01
	C. elegans	0.37/0.93/1.35	0.50/0.03		$p_r=0.27$	0.72/1.00/0.47	0.20/0.01
	E. coli	0.45/0.89/1.00	0.50/0.04		$p_r=0.30$	0.73/1.00/0.46	0.19/0.02
	H. pylori	0.45/0.83/0.89	0.59/0.08	BASF	w=2	0.71/0.71/0.00	0.50/0.02
	M. leprae	0.51/0.88/0.77	0.47/0.04		$w=3$	{1.05, 0.25, −1.13}	0.25/0.01
	P. aeruginosa	0.43/0.93/1.12	0.46/0.02		$w=4$	0.77/1.00/0.39	0.17/0.00
	S. typhi	0.42/0.96/1.18	0.43/0.04		$w=5$	0.76/1.00/0.40	0.17/0.02
	T. pallidum	0.38/0.82/1.11	0.65/0.07		$w=6$	0.74/1.00/0.43	0.18/0.01
	Y. pestis	0.54/0.87/0.69	0.47/0.03		$w=7$	0.74/1.00/0.44	0.19/0.01
	C. pneumoniae	0.63/0.86/0.44	0.41/0.07		$w=8$	0.79/1.00/0.35	0.15/0.01

by a size effect. To exclude the size effect, we consider also some artificial networks, in which the same number of random edges and nodes are added to the high-confident network. Starting from the high-confident network, at each step a new node is added by connecting it with a randomly selected node in the existing network. When the size of the network reaches 1381, we add edges between randomly selected pairs of nodes until the total number of edges is 2493. The resulting network has the same numbers of edges and nodes with the low-confident network.

The randomly added edges and nodes in the artificial networks do not lead to a similar result. This comparison may prefer to support the conjecture that an exactly constructed protein interaction network behaves perfectly fractally. The deviation of the actual structure from the perfect fractal is due to the incompleteness of the databases which are continuously being updated with newly discovered physical interactions. The high-confidence may not necessarily imply high quality.

The scaling characteristics for the real world networks are listed in Table I. We present only the results of networks whose partition functions meet the scaling relation in Eq. (1) in a wide range of q, namely $\Delta q \geq 5$. Interestingly, we find that the values of H for the real world networks are in the range of $H \geq 0.41$. Taking note of the values of error estimations δH, as presented in Table I, we have $H \geq 0.41 \approx 0.5$.

The hierarchical property may be helpful in understanding the fact that $H \geq 0.5$ for the real world networks. In the present paper, we use the definition of hierarchy proposed in [15]. That is, for a hierarchical network, besides the small-world and scale-free characteristics, there exists a simple relation between the clustering coefficient C and the degree k, $C(k) \sim k^{-1}$. Our detailed calculations show that all the considered real world networks are hierarchical in this sense.

For the Watts-Strogatz small-world (WSSW) networks [2], we can construct a regular circle lattice, with each node connected with its d right-handed nearest neighbors. Each edge is rewired with probability p_r to another randomly selected node.

As for the Barabasi-Albert scale-free (BASF) networks [3], we start from several connected nodes as a seed, at each step we add a new node and w edges from the new node to different preferentially selected nodes in the existing network. The probability for a node being selected is proportional to its degree.

The results for the constructed networks are listed in Table I. The sizes of the networks are 2 000. And the parameter d is assigned 2. The values of H for the WSSW networks are in the range of 0.15–0.31. For the BASF networks, with the increase of w the small-world effect becomes more and more significant and the value of H decreases rapidly from 0.5 ($w=2$) to <0.2 ($w>3$). Hence, for the real-world networks, the hierarchy is essential for the values of H being larger than 0.5.

The values of H for the shuffled series are almost exactly 0.5. And the corrections due to the fluctuations $\Delta\tau(0)$ are neglectable. The WSSW and BASF networks with sizes 4 000, 6 000, and 8 000 have similar characteristics (not shown in Table I).

In summary, we have found self-affine fractals embedded in spectra of complex networks. For the real world networks considered in the present work, the values of the long-range correlation exponents are in the range of $H \geq 0.5$, which may be attributed to the hierarchical properties in the sense of a

dependence of clustering on the degree. This evidence may support the idea that fractals in topological structures induce the fractals in spectra of networks.

For the constructed BASF networks, which do not have box-based fractal structures, we have also found rich multifractal structures in the spectra. However, the values of H are all significantly smaller than 0.5 for networks with $w \geq 3$. There may exist a kind of scale-invariance in the topological structures rather than the box-based fractals in the constructed networks.

One paradox that may be raised is that the box-based fractal can be explained with degree-degree anticorrelations [4], while we find the positive correlations in the spectra ($H > 0.5$) for the real world networks. Because of the degree-degree anticorrelations, the nodes tend to aggregate into many small-sized structure clusters with the hubs as centers. And there exist loosely connections between the clusters. There are strong "repulsive effects" between the levels within each cluster, but the levels for different clusters may be very close or even degenerate. That is, there will appear some locals with high density of levels in the spectrum, called level clusters. We can expect $H > 0.5$ (positive correlations in spectra) for this kind of network.

While for the BASF networks with $w \geq 3$, since there are strong correlations between the hubs, the clusters centered at the hubs will merge into a small number of large-sized clusters. The strong "repulsive effects" between the levels make the so-called "clustering of levels" impossible. Consequently, the spectra are anticorrelated ($H < 0.5$).

Hence, the difference of our results with the box-based results is not necessarily a contradiction. Obviously, the relation between the self-affine behaviors of spectra and the fractal dimension based upon box-counting approaches deserves further investigation.

Network comparison is an important topic in systems biology. It can shed light on the evolution and diseases detecting by comparing cellular networks of different species or diseased and healthy cellular networks [16]. One basic task is to design node labeling independent representations of networks and circumvent the problem of graph isomorphism. Spectra analysis of complex networks may provide useful information for that purpose.

The work has been supported by the NUS Faculty Research Grant No. R-144-000-165-112/101. It was also supported in part by the National Science Foundation of China under Grants No. 70571074 and No. 10635040. We are grateful to an anonymous referee for detailed suggestions and helpful comments.

[1] D. J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999); R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, New York, 2003).

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

[3] A. -L. Barabasi and R. Albert, Science **286**, 509 (1999).

[4] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 6392 (2005); C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).

[5] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, J. Stat. Mech.: Theory Exp. (2007) P03006.

[6] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006); J. S. Kim, K.-I. Goh, G. Salvi, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **75**, 016110 (2007).

[7] H. Yang, F. Zhao, L. Qi, and B. Hu, Phys. Rev. E **69**, 066104 (2004); F. Zhao, H. Yang, and B. Wang, *ibid.* **72**, 046119 (2005); H. Yang, F. Zhao, and B. Wang, Chaos **16**, 043112 (2006).

[8] M. Kohmoto, B. Sutherland, and C. Tang, Phys. Rev. B **35**, 1020 (1987); E. L. Albuquerque and M. G. Cottam, Phys. Rep. **376**, 225 (2003).

[9] P. CH. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature (London) **399**, 461 (1999).

[10] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, J. Geophys. Res. **111**, D01106 (2005).

[11] E. Koscielny-Bunde, J. W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin, J. Hydrol. **322**, 120 (2006).

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. CH. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, Circulation **101**, e215ce220 (2000), http://physionet.org/physiobank/

[13] J. J. Han *et al.*, Nature (London) **430**, 88 (2004).

[14] http://dip.doe-mbi.ucla.edu

[15] E. Ravasz and A.-L. Barabasi, Phys. Rev. E **67**, 026112 (2003).

[16] J. P. Bagrow, E. M. Bollt, J. D. Skufca, and D. ben-Avraham, Europhys. Lett. **81**, 68004 (2008).