Core-periphery disparity in fractal behavior of complex networks

Joon-Young Moon,[1] Dongmyeong Lee,[1] Jack H. Koolen,[2] and Seunghwan Kim[1,3]

[1]*Nonlinear and Complex Systems Laboratory, Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea*

[2]*Department of Mathematics, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea*

[3]*Institute for Edge of Theoretical Science, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea*

(Received 15 January 2011; revised manuscript received 21 June 2011; published 14 September 2011)

We show that there is a disparity in fractal scaling behavior of the core and peripheral parts of empirical small-world scale-free networks. We decompose the network into a core and a periphery and measure the fractal dimension of each part separately using the box-counting method. We find that the core of small-world scale-free networks has a nonfractal structure, whereas the periphery exhibits either fractal or nonfractal scaling. The fractal dimension of the periphery is found to coincide with one for the whole network.

DOI: 10.1103/PhysRevE.84.037103 PACS number(s): 89.75.Hc, 89.75.Fb, 05.45.Df, 64.60.aq

Gathering and analyzing data of large-scale complex networks such as the World Wide Web (WWW), social and biological networks became possible in recent years with the rapid advance of information technology [1,2]. Small-world and scale-free properties have been subjects of primary interest from the onset of complex network research. The small-world property means that the average shortest path length (ASPL) $\langle l \rangle$ between pairs of nodes in the network grows at most logarithmically with respect to the total number of nodes N: $N \approx a^{\langle l \rangle / l_0}$ (l_0 is a characteristic length) [3]. The scale-free property refers to a power law in the distribution of the number of nodes with degree (the number of links) k: $P(k) \approx k^{-\gamma}$ (γ is called the degree exponent) [4]. Also, some complex networks have been found to have fractal scaling behavior if they are measured by the box-counting method [5–8]. The fractal scaling means that there exists a power-law relation between the minimum number of boxes N_B to cover the network and the size of the boxes $l_B(N_B)$: $N_B(l_B) \approx {l_B}^{D_B}$ (D_B is the fractal dimension). This finding came as a surprise for it was generally thought that complex networks have nonfractal scaling behavior, and for the small-world property yields a exponential dependence of N on $\langle l \rangle$ [5]. For regular lattices with finite fractal dimension, the behavior of cooperative models such as the disease spreading and the Ising model depends on the fractal dimension of the lattices [9–11], whereas similar models on the scale-free small-world complex networks depend on the degree exponent of the network [12].

In this paper, we investigate empirical complex networks through the box-counting method. We show that they can be decomposed into two parts: the *core* consisting of most nodes of the network with nonfractal behavior and the *periphery* with the rest of the nodes exhibiting the same behavior as the whole network, whether it is fractal or nonfractal. Since the core part includes most of the nodes, it will play an important role in many cooperative models. However, there are instances where the peripheral part may be of critical importance such as in models of disease spreading.

Previous works suggest that subnetworks of an original network may have different fractal dimensions [8,13]. We develop this idea further from the perspective of the core-periphery partition. We first employ a criterion to define a subnetwork, the *core*, from the original network, and call the network with the rest of the nodes the *periphery*. The core we obtain contains most nodes from the original network and yet has very small diameter (maximum distance between any pair of nodes) $l_{\max}$. This core is also central in terms of network distance. We then apply the original box-counting method from Ref. [5] to measure the fractal dimension of each part separately.

In order to define the fractal dimension of an object, we employ a notion of distance that satisfies metric properties [14]. In previous works the distance between two nodes u and v of a network $d(u,v)$ has been given as the number of links in a shortest path between them [15]. This notion of distance d is called the shortest-path metric, on which our analysis is based. This distance d is computed first for the original network. Next, we apply the box-counting method restricted to the nodes in the core only and then to the nodes in the periphery only. In short, we apply the method to the nodes of each part separately, without altering the original distances between nodes.

It is known that random scale-free networks with degree exponents between 2 and 3 almost surely have an ASPL $\langle l \rangle$ of $O(\log \log N)$ [16,17]. These networks have a giant subnetwork made of a large number of nodes, residing in the central part of the network [16]. This subnetwork is densely connected and has a small diameter of $O(\log \log N)$, which constitutes the *core*. At the same time, the network itself almost always has a diameter of order $\log N$; that is, there exist nodes outside the core [16], called the *periphery*. There have been studies about the core-periphery structure of empirical networks, cellular and Internet networks in particular [18–20]. We assume that empirical networks in this work also exhibit such a core-periphery structure. In the process, we show that indeed it is easy to subtract the large, dense, and central core from the empirical networks.

All the empirical networks we focus on have both the small-world and scale-free properties with a degree exponent between 2 and 3. We study four kind of networks in this paper: the WWW of Notre Dame [4] and from the Stanford WebBase project [21], cellular networks of 43 species of which results from *Escherichia coli* and *Rickettsia prowazekii* are shown [22,23], the protein interaction network of *Saccharomyces cerevisiae* [24,25], and the Internet at the autonomous system level [26]. The properties of the networks are summarized

TABLE I. The dimensions of complex networks calculated by the box-counting method. N is the number of nodes for the corresponding network, $\langle l \rangle$ the average shortest path length between nodes, $l_{\max}$ the maximum distance among all pairs of nodes, and D_B the fractal dimension. The dimension for nonfractal behavior is denoted by ∞. Note that the dimension of the whole network is estimated from the tail part of the curve, taking box-counting data with l_B larger than the value of $\langle l \rangle$ of each network for the fit. $\langle l \rangle$ of each network is chosen as the threshold l_{th} for Criterion A of the core, $\langle l \rangle$-1 for Criterion B, and $\langle l \rangle$+1 for Criterion C. In the last two columns, the standard errors from the least-square method for the power-law model $\hat{\sigma}_\varepsilon(\text{pl})$ and the exponential curve model $\hat{\sigma}_\varepsilon(\text{ex})$ for the data from the core by Criterion A and the tail part from whole network are shown.

						Criterion A		Criterion B		Criterion C					
Network	N	$\langle l \rangle$	$l_{\max}$	γ		N	$l_{\max}$	N	$l_{\max}$	N	$l_{\max}$		D_B	$\hat{\sigma}_\varepsilon(\text{pl})$	$\hat{\sigma}_\varepsilon(\text{ex})$
WWW (Notre Dame)	325729	7.2	46	2.6	Core	280007	12	247852	10	260642	9	Core	∞	1.71	0.74
					Periphery	45722	46	77877	46	65087	46	Whole	3.2	0.37	0.69
WWW (Stanford)	8929	6.6	18	2.5	Core	7706	11	8035	12	6409	8	Core	∞	0.90	0.31
					Periphery	1223	18	894	18	2520	18	Whole	4.1	0.14	0.24
Cell (*E. coli*)	2859	4.7	18	2.3	Core	2678	8	1825	6	2134	7	Core	∞	1.18	0.65
					Periphery	181	18	1034	18	725	18	Whole	3.4	0.20	0.38
Cell (*R. prowazekii*)	817	5.0	18	2.5	Core	773	10	680	8	693	8	Core	∞	0.71	0.31
					Periphery	44	18	137	18	124	18	Whole	2.9	0.18	0.28
PIN (*S. cerevisiae*)	1458	6.8	19	2.9	Core	1143	12	756	10	960	9	Core	∞	0.72	0.16
					Periphery	315	19	702	19	498	19	Whole	∞	0.22	0.14
Internet (AS level)	22963	3.8	11	2.1	Core	21278	6	21818	8	14543	5	Core	∞	1.51	0.85
					Periphery	1685	11	1135	11	8420	11	Whole	∞	0.47	0.24

in Table I. The WWW and cellular networks have fractal scaling, while the protein interaction and Internet networks have nonfractal scaling as shown in Fig. 1. Note that the measured results for nonfractal networks are well fitted with exponential curves. Such an exponential curve can be regarded as an extreme case of the power law with the infinite fractal dimension [5]. We also note that all other 41 cellular networks from Refs. [22,23] exhibit results qualitatively similar to those of the two cellular networks shown here.

Various criteria can be given to subtract the core from an empirical network. First, inspired by Ref. [17], we choose the center of the network as the node with the largest degree and define the core as nodes within a certain distance from the center (Criterion A). There are conditions to consider in choosing an appropriate threshold distance l_{th} for the core: First, the distance must be large enough to include enough nodes to the core, so that the core can be clearly analyzable by the box-counting method. Second, the distance must be small enough to leave out enough nodes to the periphery so that the outcome for the periphery and the core can be distinguished from each other. A suitable threshold distance l_{th} for our definition is in the vicinity of the ASPL $\langle l \rangle$ of each network. Usually about 80–90% of the nodes are included in the core with such a threshold. We then apply the box-counting method to measure the fractal dimension of the core and the periphery. Results for the six networks with our core-periphery definition and $\langle l \rangle$ as l_{th} are shown in Figs. 1 and 2. The result of the core always exhibits exponential behavior regardless of the fractality of the network itself. The tail of the resulting curve from the periphery always coincides with that from the whole network.

To test the robustness of our results, we also vary l_{th}, which results in the variation in the number of core nodes. The result is shown for the WWW (Notre Dame) in Fig. 2. Choosing l_{th} in the vicinity of $\langle l \rangle$ of the network, the disparity of the core-periphery is clearly shown. Although not shown, the results are the same for other networks studied: the core exhibits exponential behavior when sufficient points are included in the core.

We try two other definitions of the core. From a network, we may delete nodes with a degree less than k and choose the largest connected component as the k core. We then add to the k core the nodes within a certain distance from it (criterion

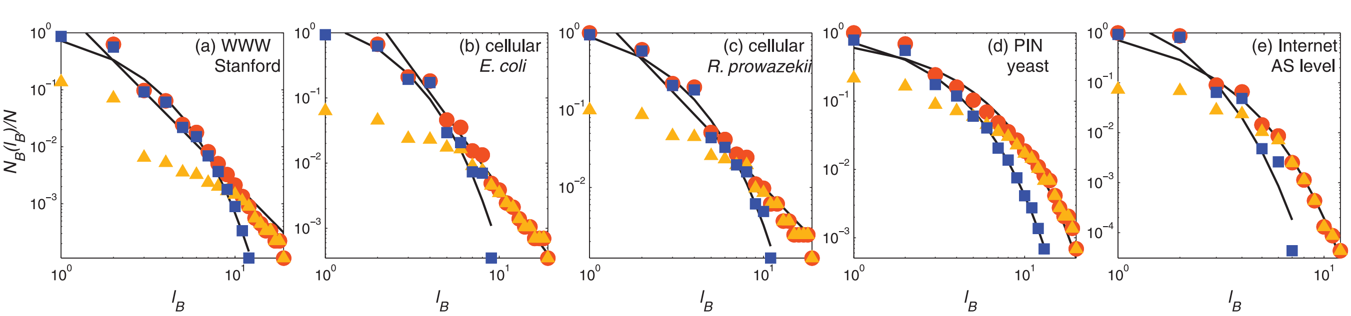

FIG. 1. (Color online) The box-counting method applied to the original networks (●, red), their core (■, blue), and their periphery (▲, orange) with two guiding lines for the core and the original network. (a) WWW from the Stanford WebBase project, (b) cellular network of *Escherichia coli*, (c) cellular network of *Rickettsia prowazekii*, (d) protein interaction network of *Saccharomyces cerevisiae*, and (e) Internet.

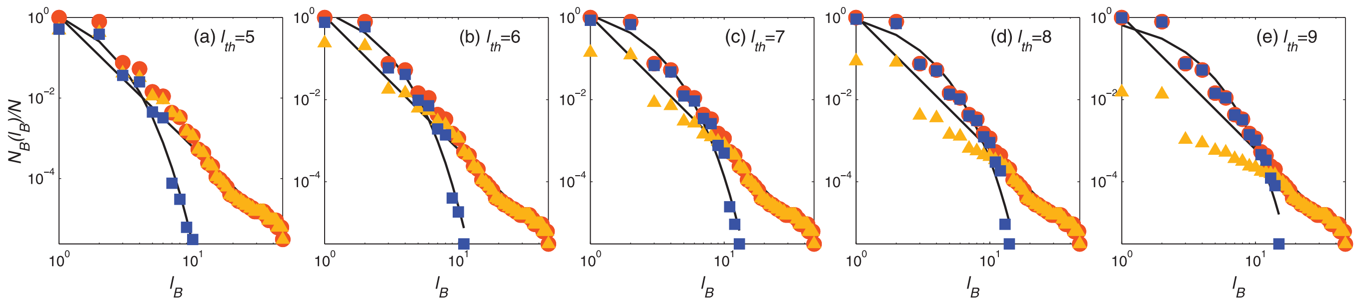

FIG. 2. (Color online) The box-counting method applied to the original network (●, red), the core (■, blue), and the periphery (▲, orange) of the WWW of Notre Dame with various threshold distances l_{th}. Two guiding lines are drawn for the core and the original network. The ASPL $\langle l \rangle$ of the network is 7.2.

B) [8,16,27]. The value of k is chosen as $k = N^{1/\log\log N}$ following Ref. [16]. Alternatively, we can choose nodes with their average distances to other nodes within a certain value and label them as the core (criterion C). The optimal l_{th} for these two alternative definitions are also in the vicinity of $\langle l \rangle$ of each network. We find that the alternative definitions yield results very similar to that from the criterion A: the core exhibits exponential behavior, while the periphery coincides with that from the whole network. The cores from different definitions overlap with each other largely, including almost all the nodes with large degrees. The results from various definitions are summarized in Table I.

We also analyze a model of the fractal network from Refs. [6,28] and obtain similar results for a plausible range of parameters. The fractal model network is built as follows: We first build a pure fractal scale-free network with fractal dimension D_B by a suitable iterative process. Then we add shortcuts between nodes with a distance r according to the probability $P(r) \sim r^{-\alpha}$ for $r > 1$. It is known that if $\alpha \lesssim 2D_B$, then the number of boxes $N_B(l_B)$ exhibits a power-law with an exponential cutoff for a large box size l_B for the box-counting method. If α is sufficiently small, the number of boxes decays exponentially. We apply our analysis to various values of α and obtain following results: (1) For the pure fractal model and the model with $\alpha \gg 2D_B$ it is difficult to distinguish the core and the periphery, because the diameter of the network $l_{\max}$ has a power-law dependence on the number of nodes N and there is no such core containing most of the nodes within a small diameter. (2) For $\alpha \lesssim 2D_B$, the analysis begins to yield results similar to those from the empirical networks. The network starts to have shortcuts connecting nodes originally far away from each other, and the core-periphery structure begins to emerge. The behavior at the tail part for peripheral nodes again coincides with that of the whole network. (3) For α sufficiently small, the network shows a small-world behavior with a small diameter. Again, it becomes difficult to distinguish the core from the periphery, because the whole network essentially collapse into a densely connected mass. The result is shown in Fig. 3. These results suggest that, whenever there a dense core and a sparse periphery exist in a complex network, the periphery contributes to the behavior of the whole network at the tail part when measured by the box-counting method.

We use a least-square method to estimate the dimension D_B from the box-counting data [29]. When testing the data for a power law (or an exponential curve), we plot the box-counting data in doubly logarithmic axes (or log-linear axes). We draw a model line fitted to the data which minimizes the sum of squared errors $S = \sum_i (y_i - \hat{y}_i)^2$. Here $\hat{y}_i$ is the predicted model value and y_i the measured one, each corresponding to $\log(\hat{N}_B(l_B)_i)$ and $\log(N_B(l_B)_i)$. We test the box-counting data from the core and the tail part from the whole network for the power-law and exponential curves. The standard error $\hat{\sigma}_\varepsilon = \sqrt{S/(i_{\max} - 2)}$ for each fit, where $i_{\max}$ is the number of different sizes of boxes l_B in each box-counting datum, is shown in the last two columns of Table I. The data from the core are always better fitted by an exponential curve, while the data from the whole network is better fitted by a power-law (exponential) for fractal (nonfractal) networks. Note that there exists an inherent difficulty in fitting the data, because the measurement itself is made over only about one order of magnitude due to the small diameter of complex networks. The best approach under the circumstance is to compare the errors for candidate fits.

We have shown that the empirical scale-free small-world complex networks are spatially inhomogeneous in their fractal behaviors. The cores of the networks exhibit nonfractal behavior. These cores include a very large number of nodes, including almost all the nodes with large degrees, which influences the outcome of cooperative models on complex

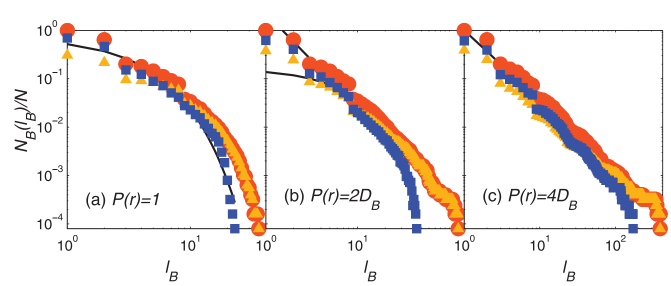

FIG. 3. (Color online) The box-counting method applied to the model networks from Refs. [6,28]. The results for the whole network (●, red), the core (■, blue), and the periphery (▲, orange) are shown. Criterion C with $l_{\mathrm{th}} = \langle l \rangle + 1$ is used in the analysis. The parameters of the model are $m = 2$ and $i = 1$, with the generation $g = 6$, and D_B is 1.46. $P(r)$ is given as (a) 1.0, (b) $2D_B$, and (c) $4D_B$, respectively. Ten percent of entire nodes are given shortcuts in each cases. Exponential guidelines are drawn for the cores in panels (a) and (b), and power-law guidelines are drawn for the whole networks in panels (b) and (c).

networks significantly. Yet the core only has a diameter of $O(\log\log N)$, meaning that most nodes of the network are packed into a small space in comparison with the diameter of the whole network, $\Theta(\log N)$. Only a few nodes reside on the spacious and sparse space between the diameter of $O(\log\log N)$ and $\Theta(\log N)$, but their spatial distribution gives rise to the fractal-like behavior of the entire network.

Though small in number, the peripheral nodes can play an important role in cooperative models on the network. Consider a simple disease-spreading process as in Ref. [3]. The process begins with a single infected node at $t = 0$. Infected nodes are removed permanently (either by immunity or death) after one unit of dimensionless time. During this time, each infected node will infect each of its healthy neighbors with probability r. After sufficient time, the disease will have either infected all the nodes or died out having infected some fraction of the nodes in the process. The core will play an important role as it contains most nodes. For example, the critical infectiousness rate r_{half} at which the disease infects half of the nodes will depend on the core heavily. Still, there are instances where the periphery plays an important role. With r high enough such that all the nodes would be infected in due time, the time T it takes to infect *all* the nodes depends on the diameter of the whole network and therefore on the diameter of the periphery. Most of the time will be spent on infecting the peripheral nodes. The role of the fractal structure of the periphery for such interacting processes on complex networks would be an interesting subject of further studies.

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MEST) (Grant No. 2010-0918849).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[3] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[5] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[6] C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).
[7] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006).
[8] J. S. Kim *et al.*, New J. Phys. **9**, 177 (2007).
[9] J. A. Redinz, J. Phys. A **31**, 6921 (1998).
[10] R. K. Pathria, *Statistical Mechanics* (Butterworth-Heinemann, Burlington, MA, 1996).
[11] J. Marro and R. Dickman, *Nonequilibrium Phase Transition in Lattice Models* (Cambridge University Press, Cambridge, UK, 1999).
[12] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[13] J. Shao *et al.*, Europhys. Lett. **84**, 48004 (2008).
[14] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications* (Wiley, New York, 1990).
[15] F. Buckley and F. Harary, *Distance in Graphs* (Addison-Wesley, Reading, MA, 1990).
[16] F. Chung and L. Lu, Proc. Natl. Acad. Sci. USA **99**, 15879 (2002).
[17] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[18] P. Holme, Phys. Rev. E **72**, 046111 (2005).
[19] S. Carmi *et al.*, Proc. Natl. Acad. Sci. USA **104**, 11150 (2007).
[20] M. da Silva, H. Ma, and A.-P. Zeng, Proc. IEEE **96**, 1411 (2008).
[21] Partial WebBase 2001 crawl [http://www-diglib.stanford.edu:8091/~testbed/doc2/WebBase/].
[22] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).
[23] R. Overbeek *et al.*, Nucleic Acids Res. **28**, 123 (2000).
[24] I. Xenarios *et al.*, Nucleic Acids Res. **28**, 289 (2000).
[25] Database of Interacting Proteins [http://dip.doe-mbi.ucla.edu/dip/].
[26] Data of July 22, 2006 [http://routeviews.org/].
[27] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).
[28] H. D. Rozenfeld, C. Song, and H. A. Makse, Phys. Rev. Lett. **104**, 025701 (2010).
[29] R. Walpole *et al.*, *Probability and Statistics for Engineers and Scientists* (Prentice Hall, Upper Saddle River, NJ, 2002).