Journal of the Korean Physical Society, Vol. 56, No. 3, March 2010, pp. 1020∼1024

# Fractal Network in the Protein Interaction Network Model

Pureun Kim and Byungnam Kahng*
*Department of Physics and Astronomy, Seoul National University, Seoul 151-747*

(Received 22 September 2009)

Fractal complex networks (FCNs) have been observed in a diverse range of networks from the World Wide Web to biological networks. However, few stochastic models to generate FCNs have been introduced so far. Here, we simulate a protein-protein interaction network model, finding that FCNs can be generated near the percolation threshold. The number of boxes needed to cover the network exhibits a heavy-tailed distribution. Its skeleton, a spanning tree based on the edge betweenness centrality, is a scaffold of the original network and turns out to be a critical branching tree. Thus, the model network is a fractal at the percolation threshold.

PACS numbers: 68.37.Ef, 82.20.-w, 68.43.-h
Keywords: Fractal complex network, Percolation, Protein interaction network
DOI: 10.3938/jkps.56.1020

## I. INTRODUCTION

Fractal complex networks (FCNs) have been discovered in diverse real-world systems [1, 2]. Examples include the co-authorship network [3], metabolic networks [4], the protein interaction networks [5], the World-Wide Web [6] and so on. Here, FCNs are the networks satisfying the fractal scaling [7],

$$N_B(\ell_B) \sim \ell_B^{-d_B}, \tag{1}$$

where $N_B$ is the number of boxes needed to cover the entire network and $\ell_B$ is the box size. While such FCNs are ubiquitous, there exist only a few FCN models [2,8]. Here, we will introduce a stochastic model to generate FCNs using protein interaction network model.

It was viewed that the fractal scaling originates from the disassortative correlation between degrees of two neighboring nodes [9]. Thus, hubs are not directly connected to one another [10,11]. Based on this property, a toy model and hierarchical models were introduced [2]. On the other hand, FCNs are regarded as the composition of a skeleton and shortcuts. The skeleton is a spanning tree of the underlying network based on the edge-betweenness centrality [12, 13] or load [14], which can be regarded as the communication backbone of underlying network [15]. Since the skeleton is composed preferentially of high betweenness edges, edges connecting different modules may be well represented, preserving overall modular structure. In fact, FCNs are modular networks in which hubs are central nodes of each module and separated from one another [3]. This picture is consistent with that of the hub-repulsion model [2]. The fractal scaling of a FCN originates from the fractality of its skeleton underneath it [8]. The skeleton is regarded as a critical branching tree: It exhibits a plateau in the mean branching number function $\bar{n}(d)$, defined as the average number of offsprings created by nodes at a distance $d$ from the root. Random branching tree exhibits such a plateau, the average value of which we denote as $\bar{n}$. Such a persistent branching structure underlies the fractality of the skeleton, as it is known that the random branching tree is a fractal for the critical case, namely, $\bar{n} = 1$ [16]. Specifically, SF random branching tree with the branching probability $b_n$ that each branching event produces $n$ offsprings, $b_n \sim n^{-\gamma}$, generates a SF tree, which is a fractal when $\bar{n} = 1$. Using this property, a fractal network model was introduced [8].

Recently, a fractal complex network was observed in the intermediate regime in the evolution of co-authorship networks [3]. Hinted from the evolution pattern, a stochastic model was introduced, which exhibits a percolation transition. FCNs can be generated near the percolation threshold. Similar to this behavior, here we study the fractality in a protein interaction network model introduced by Solé *et al.* [17], called the Solé model hereafter.

Protein interaction networks (PINs) have been studied in a variety of organisms including viruses [18], yeast [19], *C.elegans* [20], *etc.* Previous studies have focused on dynamical or computational aspects of interacting proteins as well as their potential links. Moreover, it has been of interest to construct an evolution model of the PIN with biological relevant ingredients. One of successful models is the one introduced by Solé *et al.* In this model, they used three edge dynamics, duplication, mutation, and

*E-mail: bkahng@snu.ac.kr; Fax: +82-2-884-3002

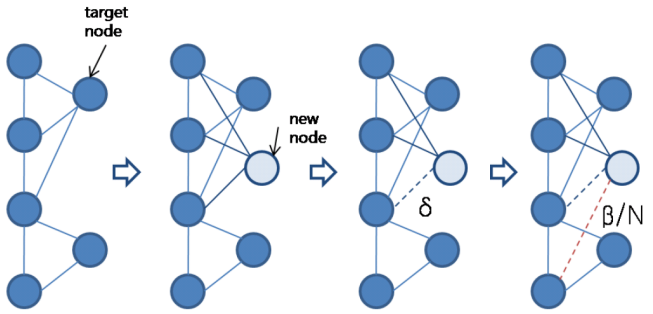


Fig. 1. The protein-protein interaction network model introduced by Solé *et al.* [17]

divergence, as key ingredients of the evolution of protein interactions. Similar models follow this model, which are listed in the references [21–25]. In this paper, we used the Solé model to construct FCNs, because analytic solution of the perolation transition is known for the Solé model. We find that as in the case of the co-authorship network [3], indeed FCNs can be obtained near the percolation threshold.

Complex network becomes a fractal if it is composed of well-organized modules within it, in which many short range edges and few long range edges are contained [26–29]. Then, why do PINs have few long range edges? Because, proteins within a particular module are the one playing a particular function. For example, proteins which take part in cell cycle have to be closely connected, but they do not need to connect to proteins which have different functions such as signal transduction. In this model, modules are made by short range connections such as duplication and divergence. Furthermore, few densities of long range connections, that is, few mutation, make this model fractal.

## II. THE SOLÉ MODEL FOR PIN

Solé *et al.* [17] proposed a protein-protein interaction network model which contains three ingredients: (i) duplication, (ii) divergence, and (iii) mutation. The model is a growing network model in which a node (protein) is created in the system at each time step. This is achieved in the form of duplication: The new node duplicates a randomly chosen pre-existing protein and its links are also endowed from its ancestor. Among those links, links of the new protein are deleted with probability $\delta$ (divergence) and each new protein also links to any pre-existing node with probability $\beta/N$ (mutation), where $N$ is the total number of nodes at each time step. The two parameters $\delta$ and $\beta$ control the densities of the short-ranged and the long-ranged edges, respectively. This model has been solved analytically [30]. Here, we review some important analytic results relevant to our works in this paper.

Let $n_s(N)$ be the number of $s$-size clusters per node at time step $N$, and $g(z) = \sum_s n_s z^s$ the generation function for $n_s$, where the sum excludes the giant percolating cluster. $g(1)$ is the fraction of finite clusters and $g'(1)$ is the average cluster size, *i.e.*, $\langle s\rangle = \sum s^2 n_s$. The model exhibits unconventional percolation transition in which the parameter $\delta$ turns out to be irrelevant, and thus it is ignored for the time being. Within this scheme, the analytic solution yields that

$$\langle s\rangle = \begin{cases} \frac{1-2\beta-\sqrt{1-4\beta}}{2\beta^2} & \text{for } \beta \le \beta_c, \\ \frac{e^{-\beta G}+G-1}{\beta(1-e^{-\beta G})} & \text{for } \beta > \beta_c, \end{cases} \tag{2}$$

where the size of the giant cluster $G = 1 - g(1) = 1 - \sum_s s n_s$. $\beta_c$ is the percolation threshold and obtained as $\beta_c = 1/4$. The cluster-size distribution follows a power law,

$$n_s \sim s^{-\tau}, \tag{3}$$

where the exponent is solved as

$$\tau = 1 + \frac{2}{1-\sqrt{1-4\beta}}. \tag{4}$$

This power-law behavior holds in the entire range $\beta < \beta_c$ in contrast to the behavior of the conventional percolation transition. At the transition point $\beta = \beta_c$, the cluster-size distribution decays as $n_s \sim 1/[s^3(\ln s)^3]$.

The order parameter of the percolation transition is written as

$$G(\beta) \propto \exp\Big(-\frac{\pi}{\sqrt{4\beta-1}}\Big). \tag{5}$$

Thus, all derivatives of $G(\beta)$ vanishes as $\beta \to \beta_c$, and the transition is of infinite order.

The degree distribution was also studied in [30] for general $\delta$. When $\delta > 1/2$ and $\beta > 0$, the degree distribution follows a power law $P_d(k) \sim k^{-\gamma}$, where the degree exponent is determined from the relation,

$$\gamma(\delta) = 1 + \frac{1}{1-\delta} - (1-\delta)^{\gamma-2}. \tag{6}$$

## III. SIMULATION RESULTS

We performed extensive numerical simulations for the Solé model with system sizes $N = 10^3$ and $10^4$ and various parameter values $\beta$ and $\delta$. The obtained results are as follows:

First, to see a percolation transition behavior, we measure the mean component size $\langle s\rangle$ as a function of $\beta$ at a fixed parameter value $\delta = 0.95$, corresponding to the case with little duplication. We find that the mean component size exhibits a peak at a point $\beta_c$, which is estimated to be $\beta_c \approx 0.29$ (Fig. 2). This critical point is regarded as a percolation threshold. The percolation threshold $\beta_c$ varies as a function of $\delta$. Thus, we plot in Fig. 3 the mean component size as a function of $\beta$ and $\delta$. The peak

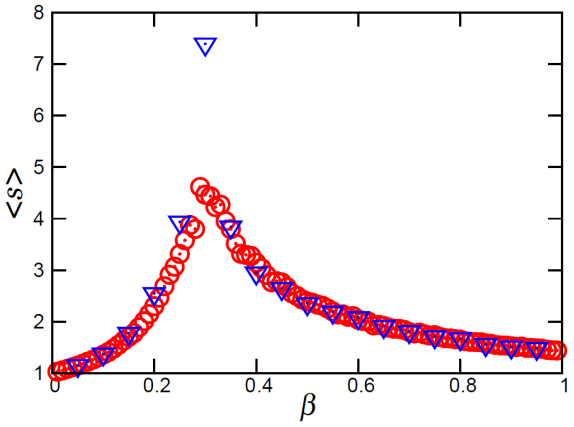


Fig. 2. Mean component size $\langle s \rangle$ versus $\beta$ at $\delta = 0.95$. Data, obtained from system sizes $N = 10^4$ ($\triangledown$) and $10^5$ ($\circ$), display peaks at the percolation transition, which is to be $\beta \approx 0.29$. All data points are averaged over 100 configurations.

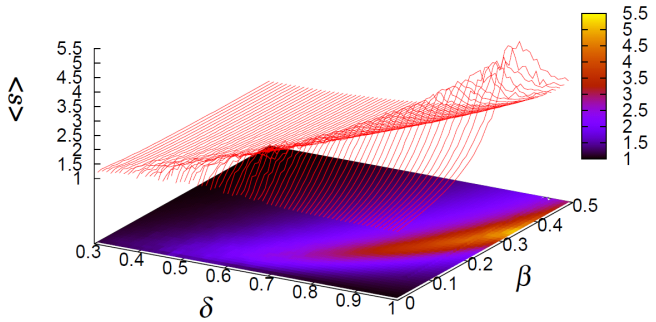


Fig. 3. Mean component size $\langle s \rangle$ as a function of two parameters $\beta$ and $\delta$ for $N = 10^4$. The peak line is the boundary between two phases, the percolating and the non-percolating phases.

locus lies in the small $\beta$ region, indicating that a small fraction of long-range edges are sufficient to develop the giant component.

Second, we show a giant component of the model network with small system size $N = 10^4$ at the percolation threshold in Fig. 4. This network is constructed with parameter values $\beta = 0.29$ and $\delta = 0.95$. The network topology is effectively a tree but with small-size loops within it.

Third, we examine the degree distribution of the giant component at evolution steps $N = 10^3$ and $N = 10^4$ with parameter values used in Solé *et al.* [17] $\delta = 0.58$ and $\beta = 0.16$ and show them in Fig. 5. It shows heavy-tailed behaviors, but tends to converge to a power-law behavior with increasing $N$. Solid line in Fig. 5 has a slope $-2.94$, the theoretical value obtained from Eq. (6). Thus, numerical data are expected to converge to the theoretical prediction asymptotically.

Fourth, in order to see the fractality of the model network, we measure the number of boxes $N_B$ defined in Eq. (1) as a function of box size $\ell_B$ using the box-covering method [8]. In Fig. 6, $N_B(\ell_B)$ exhibits a heavy-tailed dis-

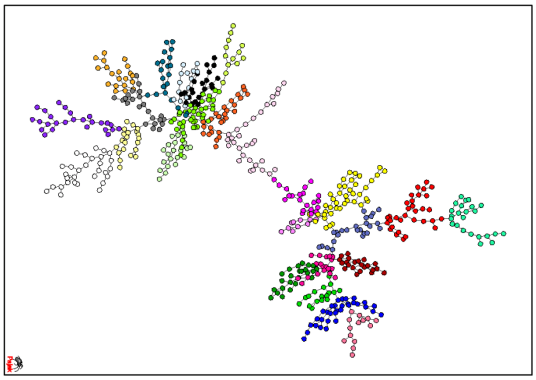

Fig. 4. Snapshot of the giant component near the percolation threshold $\delta = 0.95$ and $\beta = 0.29$ for size $N = 568$.

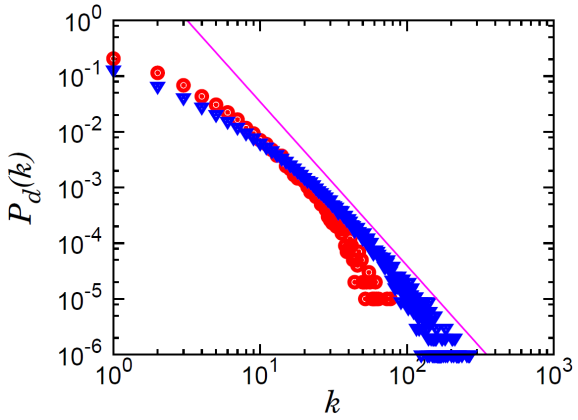


Fig. 5. Degree distribution $P_d(k)$ versus $k$ for the giant component of the Solé model with $\delta = 0.58$ and $\beta = 0.16$. Shown are the distributions for $N = 10^3$ ($\circ$) and $10^4$ ($\triangledown$). The solid line is the predicted line with slope $-2.94$.

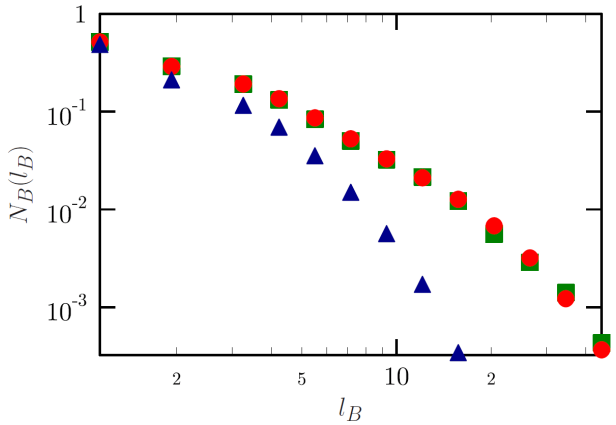


Fig. 6. Fractal scaling analysis of the giant component (◯) and its skeleton (■) near the percolation transition point $\delta = 0.95$ and $\beta = 0.29$ (◯, ■). The number of boxes follows a heavy-tailed distribution. However, data obtained at $\delta = 0.6$ and $\beta = 0.3$ (▲), located far away from the percolation threshold, decay faster than the previous one. All data points are log-binned and are averaged over 100 configurations.

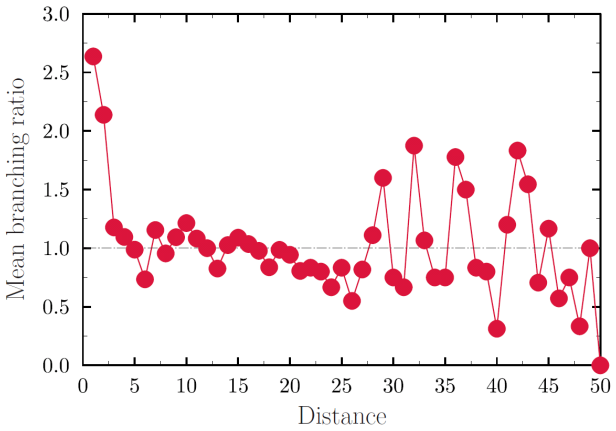


Fig. 7. Mean branching number as a function of distance from a root for the skeleton of the giant component produced by the parameter values $\delta = 0.95$ and $\beta = 0.29$.

tribution with respect to $\ell_B$ when the data are obtained near the percolation threshold ($\delta = 0.95$ and $\beta = 0.29$). The numbers of boxes covering the skeleton for each box size are also shown in Fig. 6: They overlap with those covering the entire network. Since power-law behavior is not manifest, one may wonder if this model network is indeed a fractal. Thus, we present $N_B(\ell_B)$ for the network obtained from different parameter values, particularly, locating far away from the percolation threshold. Indeed, $N_B(\ell_B)$ for this case decays fast compared with that obtained near the percolation threshold. Moreover, we show next that the skeleton is a critical branching tree with mean branching rate is one. Since the critical branching tree was shown to be a fractal analytically, and together with the above numerical results, we can claim that the original model is also a fractal.

Lastly, the criticality of the skeleton is checked. We measure the mean branching number function $\bar{n}(d)$ as a function of distance from a root. Indeed, it fluctuates around one, implying that the skeleton can be regarded as a critical branching tree and thus a fractal manifestly (Fig. 7). Since the box numbers to cover the entire network with each box size and the skeleton only are the same, we can say that the entire network is a fractal.

## IV. CONCLUSION

In summary, we simulated the protein-protein interaction network model introduced by Solé *et al.*, finding that the obtained network near the percolation threshold was a fractal. We measured the number of boxes needed to cover the network by using the box-covering method for both the full network and its skeleton and found that their values were approximately the same and fat-tailed. We also measured the mean branching ratio for the skeleton and found that it fluctuated around the critical value of one as a function of the distance from the root. The fat-tailed distribution in fractal analysis and the criticality of the skeleton with mean branching ratio being one support the fractality of the model network. Thus, we conclude that the protein-protein interaction network model obtained near the percolation threshold can be used to generate fractal complex networks.

## ACKNOWLEDGMENTS

This work is supported by KRCF and SNU-ORA.

## REFERENCES

[1] C. Song, S. Havlin and H. A. Makse, Nature **433**, 392 (2005).
[2] C. Song, S. Havlin and H. A. Makse, cond-mat/0507216 (2005).
[3] D. Lee, K.-I. Goh and B. Kahng, J. Korean Phys. Soc. **52**, 197-202 (2008).
[4] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabasi, Nature **407**, 651 (2000).
[5] I. Xenarious, Danny W. Rice, Lukasz Salwinski, Marisa K. Baron, Edward M. Marcotte and David Eisenberg, Nucleic Acids Res. **28**, 289 (2000).
[6] R. Albert, H. Jeong and A.-L. Barabasi, Nature **401**, 130 (1999).
[7] J. Feder, *Fractals* (Plenum, New York, 1988).
[8] K.-I. Goh, G. Salvi, B. Kahng and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006).
[9] S.-H. Yook, F. Radicchi and H. Meyer-Ortmanns, Phys. Rev. E **72**, 045105 (2005).
[10] J. S. Kim, B. Kahng, D. Kim and K.-I. Goh, J. Korean Phys. Soc. **52**, 350 (2008)
[11] J. S. Kim, B. Kahng and D. Kim, Phys. Rev. E **79**, 067103 (2009)
[12] L. C. Freeman, Sociometry **40**, 35 (1977).
[13] M. Girvan and M. E. J. Newman, Proc. Nat. Acad. Sci. U.S.A. **99**, 7821 (2002).
[14] K.-I. Goh, B. Kahng and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[15] D.-H. Kim, J. D. Noh and H. Jeong, Phys. Rev. E **70**, 046126 (2004).
[16] Z. Burda, J. D. Correia and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
[17] R.V. Sole, R. Pastor-Satorras, E. D. Smith and T. Kepler, Adv. Complex Syst. **5**, 43 (2002).
[18] M. Flajolet, G. Rotondo, L. Daviet, F. Bergametti, G. Inchauspe, P. Tiollais, C. Transy and P. Legrain, Gene **242**, 369 (2000).
[19] T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara and Y. Sakaki, Proc. Nat. Acad. Sci. U.S.A. **97**, 1143 (2000).
[20] A. J. M. Walhout, R. Sordella, X. W. Lu, J. L. Hartley, G. F. Temple, M. A. Brasch, N. Thierry-Mieg and M. Vidal, Science **287**, 116 (2000).
[21] J. C. Nacher, M. Hayashida and T. Akutsu, Physica A, **367**, 538 (2006).
[22] G. P. Karev, Y. I. Wolf, A. Y. Rzhetsky, F. S. Berezovskaya and E. V. Koonin, BMC Evol. Biol. **2**, 18 (2002).

[23] S. Wuchty, Mol. Biol. Evol. **18**, 1694 (2001).
[24] G. P. Karev, F. S. Berezovska and E. V. Koonin, Bioinformatics **21**, 12 (2005).
[25] A. Beyer and T. Wilhelm, Bioinformatics **21**, 1620 (2005).
[26] Jing-Dong J. Han, Nicolas Bertin, Tong Hao, Debra S. Goldberg, Gabriel F. Berriz, Lan V. Zhang, Denis Dupuy, Albertha J. M. Walhout, Michael E. Cusick, Frederick P. Roth and Marc Vidal, Nature **430**, 88 (2004).
[27] Jose B. Pereira-Leal, Anton J. Enright and Christos A. Ouzounis, Bioinformatics **54**, 49 (2004).
[28] Victor Spirin and Leonid A. Mirny, Proc. Nat. Acad. Sci. U.S.A. **100**, 12123 (2003).
[29] Jingchun Chen and Bo Yuan, Bioinformatics **22**, 2283 (2006).
[30] J. Kim, P. L. Krapivsky, B. Kahng and S. Redner, Phys. Rev. E **66**, 055101 (2002).