

On General Laws of Complex Networks

Wenjun Xiao¹, Limin Peng², and Behrooz Parhami³

¹ School of Software Engineering, South China University of Technology,
Guangzhou 510641, P.R. China
wjxiao@scut.edu.cn

² Department of Computer Science, South China University of Technology,
Guangzhou 510641, P.R. China
penglm86@126.com

³ Dept. Electrical & Computer Eng., University of California,
Santa Barbara, CA 93106-9560, USA
parhami@ece.ucsb.edu

Abstract. By introducing and analyzing a renormalization procedure, Song et al. [1] draw the conclusion that many complex networks exhibit self-repeating patterns on all length scales. First, we aim to demonstrate that the aforementioned conclusion is inadequately justified, mainly because their equation (7) on the invariance of degree distribution under renormalization does not hold in general. Secondly, Barabási and Albert [2] find that many large networks exhibit a scale-free power-law distribution of vertex degrees. They show this common feature to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to those that are already well connected. We show that when vertex degrees of large networks follow a scale-free power-law distribution with the exponent $\gamma \geq 2$, the number of degree-1 vertices, when nonzero, is of the same order as the network size N and that the average degree is of order less than $\log N$. Given that many real networks satisfy these two conditions, our results add another necessary characteristic of the scale-free power-law distribution of vertex degrees in such networks. Our method has the benefit of relying on conditions that are static and easily verified. They are verified by many experimental results of diverse real networks.

Keywords: complex network, self-similarity, scale-free, computer network, computer communication.

1 Introduction

Complex systems with many components and associated interactions arise in nature, society, and many human artifacts. Interactions in such systems can be modeled by networks composed of vertices and links, which are in turn abstracted as undirected or directed graphs. A graph G , denoted as $G = (V, E)$, has a set V of vertices or nodes and a set of edges or links, where each edge is defined by a pair of vertices (ordered pair, for directed graphs) [7]. Complex systems in the three categories of natural, societal, and synthetic include:

1. Protein interactions, metabolic systems, contagious diseases
2. Acquaintances, movie-actor peer group, research collaborators
3. Power grid, Internet connectivity, Worldwide Web linkages.

Two models of complex networks have been studied extensively [1-10]: the small-world model and the scale-free one. The small-world model features localized clusters that are connected by occasional long-range links, leading to an average distance between vertices that grows logarithmically with the network size N . Watts and Strogatz [3] investigated mechanisms via which a regular network can be transformed into a small-world network, without significantly modifying the vertex-degree distribution, and quantified the parameters that characterize the resulting structures.

Scale-free networks, on the other hand, tend to have uneven vertex connectivities, so that a certain fraction of vertices, independent of network size, are highly connected (the hubs). Barabási and Albert [2] demonstrated that the scale-free power-law distribution of vertex degrees in many large networks is a direct consequence of two generic mechanisms that govern network formation: (i) Networks expand over time through the addition of new vertices, and (ii) New vertices attach preferentially to those that are already well connected. It is well-known that scale-freeness of a network has significant implications for its diffusion properties and its robustness.

In this paper, we focus on scale-free networks. After reviewing the parameters and key attributes of such networks in Section 2, we provide necessary characteristic conditions for scale-free complex networks and show that these conditions are both easy to verify and satisfied by many natural and man-made scale-free networks (Section 3). Section 4 contains our conclusions and some directions for further research.

2 On Self-similarity of Complex Networks

We begin by reviewing some relevant properties of complex networks [1-6]. Two models of real complex networks have been studied extensively: the small-world model and the scale-free one. We shall focus first on scale-free networks. To avoid confusion with equations in this paper, equation numbers in our references will be enclosed in square brackets.

For many complex networks, the probability distribution $P(k)$ of the number of degree- k vertices, also known as the degree distribution, can be represented (independent of scale) by a power law with characteristic exponent γ ([2] means that this is equation [2] in ref. 1):

$$P(k) \approx k^{-\gamma}. \quad [2] (1)$$

The renormalization method, introduced in ref. 1, is as follows. Let G be a network of degree distribution $P(k)$ satisfying equation (1). Vertices of G are covered by N_B boxes of linear size l_B . Boxes are then viewed as vertices of a renormalized network G' , with two such vertices (boxes) connected in G' if and only if there exists at least one link between their constituent vertices in G . Song et al. [1] guided only by experimental results for WWW, then proceed with the assumption that the degree distribution is invariant under renormalization:

$$P(k) \rightarrow P(k') \approx (k')^{-\gamma}. \quad [7] (2)$$

And here lies the problem: equation (2) is neither proved mathematically nor verified adequately by experimentation with different complex networks. Hence, the validity of equation (2) is suspect. We shall endeavor to show equation (2) invalid by examining both the theoretical and experimental analyses of Song et al [1].

In the theoretical analysis that follows, we refer to, and use, four equations from ref. 1, reproduced below for completeness. Here the network G is called self-similarity if the equation (3) holds.

$$\langle M_B(l_B) \rangle \equiv N/N_B(l_B) \approx l_B^{d_B}. \quad [5] (3)$$

$$k \rightarrow k' = s(l_B)k. \quad [6] (4)$$

$$s(l_B) \approx l_B^{-d_k}. \quad [8] (5)$$

$$\gamma = 1 + d_B/d_k. \quad [9] (6)$$

In the equations above, $\langle M_B(l_B) \rangle$ is the average mass of (or the number of vertices in) a box, d_B is the fractal dimension (or box dimension) derived from $N_B \approx l_B^{-d_B}$, $s(l_B) < 1$ is the scaling of vertex degrees owing to renormalization (i.e., the ratio k'/k), and d_k is a new exponent characterizing the variation of s with l_B .

We prove that equations (4)-(6) can be derived directly from equation (3), without using the suspect equation (2). For this, we set $s = (N/N')^{1/(1-\gamma)}$, where N' is the size of G' . By equation (3), we have $s \approx l_B^{d_B/(1-\gamma)}$. Let $d_k = d_B/(\gamma-1)$, we obtain equation (6), which leads to $s \approx l_B^{-d_k}$; viz., equation (5). Finally, we obtain equation (4) from setting $k' = s k$, complete our proof. However, we cannot obtain equation (2) from equation (3). On the other hand, it is easily verified that equation (2) is equivalent to $n'(k') \approx n(k)$, where $n(k)$ and $n'(k')$ represent the number of vertices of degrees k and k' in networks G and G' , respectively.

It is easily shown that multiple renormalizations at fixed l_B can be achieved by an equivalent renormalization. In fact, defining

$$N^{(i)}/N^{(i+1)} \approx l_B^{d_B^{(i)}}. \quad (7)$$

for $i = 0, 1, \dots, t$ and $N = N^{(0)}$, we have

$$N/N^{(t)} \approx l_B^{d_B^{(0)} + \dots + d_B^{(t-1)}}. \quad (8)$$

Next, setting $s^{(i)} = (N^{(i)}/N^{(i+1)})^{1/(1-\gamma)}$, we obtain $s^{(i)} \approx l_B^{d_B^{(i)}/(1-\gamma)}$. Finally, from $d_k^{(i)} = d_B^{(i)}/(\gamma-1)$, we get $s^{(i)} \approx l_B^{-d_k^{(i)}}$. Taking $k^{(0)} = k$ and $k^{(i+1)} = s^{(i)}k^{(i)}$ leads to the following two equations:

$$k^{(t)} = s^{(0)} \dots s^{(t-1)} k. \tag{9}$$

$$s^{(0)} \dots s^{(t-1)} \approx l_B^{-d_k^{(0)} - \dots - d_k^{(t-1)}}. \tag{10}$$

We have thus demonstrated that the hypothesis which Song et al. [1] derive experimentally (namely, that based on Fig. 2d of ref. 1, the degree distribution of the WWW of special sizes is invariant under renormalization for different box sizes) is not supported by theoretical analysis, leading to serious doubts regarding its correctness in general. As a consequence, it is not surprising that many complex networks (Internet, protein interactions, and some random networks) lack self-similarity, as indicated in the supplementary materials of ref. 1. This can be explained theoretically as follows. By equation (7), we may set

$$N^{(i)} / N^{(i+1)} \approx C_i l_B^{d_B^{(i)}}. \tag{11}$$

thus, leading to

$$N / N^{(t)} \approx C_0 \dots C_{t-1} l_B^{d_B^{(0)} + \dots + d_B^{(t-1)}}. \tag{12}$$

In equation (12), the product $C_0 \dots C_{t-1}$ may be an exponential function of the average distance \bar{l} when $C_i > 1$ and $t \approx \bar{l}$. This would imply that the network size is an exponential function of the average distance \bar{l} , thus, the network lacks self-similarity in this case.

3 Necessary Conditions for Scale-Free Networks

In the following, we study the conditions for vertex degrees of complex networks having scale-free power-law distribution. We assume that the network is connected; similar arguments apply to disconnected networks. Let $P(k)$ be the probability distribution of the number of vertices of degree k , as previously defined. Let A denote the average vertex degree and n_k the number of degree- k vertices. We have $M = \frac{1}{2}NA$ and $n_k = NP(k)$, where N and M are numbers of vertices and edges, respectively. The preceding definitions imply

$$\sum_{k=1}^{N-1} n_k = N. \tag{13a}$$

$$\sum_{k=1}^{N-1} kn_k = 2M. \tag{13b}$$

Supposing that $n_1 \neq 0$, we have $n_k = n_1 P(k) / P(1)$ and

$$\sum_{k=1}^{N-1} [P(k) / P(1)] = N / n_1. \tag{14a}$$

$$\sum_{k=1}^{N-1} [k P(k) / P(1)] = 2M / n_1. \tag{14b}$$

For scale-free networks, we have $P(k) = P(1) k^{-\gamma}$, which leads to $\sum_{k=1}^{N-1} k^{-\gamma} = N/n_1$. Therefore, assuming $\gamma \geq 2$, which is known to hold for many scale-free networks [4,5], we have

$$N/n_1 \leq \sum_{k=1}^{\infty} k^{-\gamma} \leq \sum_{k=1}^{\infty} k^{-2} = \pi^2/6. \tag{15}$$

This leads to the conclusion $N \approx n_1$. Let $f(k) = P(k)/P(1)$. When $N \approx n_1$, equations (14) yield

$$\sum_{k=1}^{N-1} f(k) \approx 1. \tag{16a}$$

$$\sum_{k=1}^{N-1} k f(k) \approx A. \tag{16b}$$

Assuming that all logarithms are in base 2, equations (16) yield $1 < A \leq \log N$, given that $\sum_{k=1}^N k^{-1} \approx \log N$. Thus, we have proved that for many real complex networks of scale-free power-law distribution with $\gamma \geq 2$, the number of degree-1 vertices, when nonzero, is of the same order as the network size N and that the average degree is of order less than $\log N$.

On the other hand, if $\log \log N \leq A \leq \log N$, then equations (16) imply

$$f(k) \approx k^{-\gamma}. \tag{17}$$

We now elaborate on equation (17). There are three canonical cases for the function $f(k)$ when $\log \log N \leq A \leq \log N$: (a) $f(1) = 1$ and $f(k)$ a constant for $k \neq 1$; (b) $f(k) = e^{1-k}$; and (c) $f(k) \approx k^{-\gamma}$. For case (c), equations (16) hold when $2 \leq \gamma < 3$. However, for cases (a) and (b), equation (16b) is not satisfied if equation (16a) holds. Note that more complex functions satisfying equations (16) exist. For example, one can define the function

$$f(k) \approx e^{1-k}, \text{ if } k \leq \log N; f(k) \approx k^{-\gamma}, \text{ otherwise.} \tag{18}$$

which satisfies equations (16). As a consequence of our results above, the scale-free property of complex networks must be viewed as an approximate or fuzzy property.

The preceding leads to a model for scale-free networks, known to satisfy $k^\gamma = n_1/n_k$. Taking this equality to be exact, and noting that the right-hand side is a rational number, we can readily prove that k^γ must be an integer that divides n_1 . Then, n_1 must be divisible by the least common multiple of $k_1^\gamma, k_2^\gamma, \dots, k_l^\gamma$, where $1 = k_1 < k_2 < \dots < k_l$ is the degree sequence of the network. This is a general model for scale-free networks that we now aim to study further.

Let $n_1 = c[k_1^\gamma, k_2^\gamma, \dots, k_l^\gamma] \neq 0$, where c is constant. If $\gamma \geq 2$ then

$$N = n_1 \sum_{i=1}^l k_i^{-\gamma} \leq \frac{\pi^2}{6} n_1 < \frac{5}{3} n_1 \text{ and } M = \frac{1}{2} n_1 \sum_{i=1}^l k_i^{1-\gamma} \leq \frac{1}{2} n_1 \leq N \log \log N.$$

Now we give a sufficient and necessary condition that a connected graph is a tree.

Proposition. Assume that G is connected and scale-free, then G is a tree

$$\Leftrightarrow \sum_k k^{1-\gamma} = \frac{2N-2}{n_1}.$$

Proof. G is a tree $\Leftrightarrow 2N-2 = N \sum_k P(1)k^{1-\gamma} = N \sum_k \frac{n_k}{N}k^{1-\gamma} = n_1 \sum_k k^{1-\gamma}$.

Table 1. Number of vertices (N), number of edges (M), average degree (A), and characteristic exponent (γ) in some complex networks

Network	N	M	A	γ
Internet	10,687	31,992	5.98	2.5
Film actors	449,913	25,516,482	113.43	2.3
Metabolic network	765	3686	9.64	2.2
Protein interactions	2115	2240	2.12	2.4

Table 1 lists the parameters N , M , A , and γ for several real scale-free networks [4,5]. We note that $A \approx \frac{1}{2} \log N$ (respectively, $5 \log N$, $\log N$, and $\log \log N$) for the Internet (film actors, metabolic, and protein interaction) network.

4 Conclusion

We have shown that for many real networks of scale-free power-law degree distribution with the exponent $\gamma \geq 2$, the number of degree-1 vertices, when nonzero, is of the same order as the network size N and that the average degree is of order less than $\log N$. Our method has the benefit of relying on conditions that are static and easily verified for any network. Such distributions are known to be applicable to diverse fields of study, including computer communication and software architecture. However it is worth to further research networks of scale-free power-law degree distribution with the exponent $\gamma < 2$.

Such extensions and variations will further broaden the applications of our results in diverse subfields within computing, communication, biology, and the social sciences.

Acknowledgements. Research of W. Xiao is supported by Guangdong Key Laboratory of Computer Network (CCNL200705) and Laboratory of Basic Software and Applied Construction Technology of Guangdong(2006B80407001).

References

1. Song, C., Havlin, S., Makse, H.A.: Self-similarity of complex networks. Nature 433, 392–395 (2005)
2. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)

3. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* 393, 440–442 (1998)
4. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97 (2002)
5. Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* 45, 167–256 (2003)
6. Xiao, W.J., Parhami, B.: Cayley Graphs as Models of Deterministic Small-World Networks. *Information Processing Letters* 97, 115–117 (2006)
7. Biggs, N.: *Algebraic Graph Theory*. Cambridge University Press, Cambridge (1993)
8. Myers, C.R.: Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. *Physical Review E* 68, 046116 (2003)
9. Montoya, J.M., Sole, R.V.: Small World Patterns in Food Webs. *Journal of Theoretical Biology* 214, 405–412 (2002)