On Necessary Conditions for Scale-Freedom in Complex Networks with Applications to Computer Communication Systems[1]

Wenjun Xiao[1], Weidong Chen[1], Wenhong Wei[1], Brooze Parhami[2]
1. *Department of Computer Science, South China University of Technology, Guangzhou, 510641, P.R.China*
2. *Department of Electrical & Computer Eng., University of California, Santa Barbara, CA 93106-9560, USA*
wjxiao@scut.edu.cn

Abstract

Many complex networks exhibit a scale-free, power-law distribution of vertex degrees. This common feature is a consequence of two generic mechanisms relating to the formation of real networks: (i) Networks tend to expand over time through the addition of new vertices, and (ii) New vertices attach preferentially to those that are already well connected. We show that for many natural or man-made complex networks possessing a scale-free power-law distribution with the exponent $\gamma > 2$*, 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 log* N*. Our results expose another necessary characteristic of such networks. Furthermore, our method has the benefit of relying only on conditions that are static and easily verified for arbitrary networks. We use the preceding results to derive a closed-form formula approximating the distance distribution in scale-free networks. Such distributions find extensive applications in computer communication networks and software architecture.*

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 E of edges or links, where each edge is defined by a pair of vertices (ordered pair, for directed graphs). 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 [1-4]: 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 [2] 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 [1] 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-freedom 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 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

[1]This Work is supported by Guangdong Key Laboratory of Computer Network (CCNL200705), Laboratory of Software and Applied Technology of Guangdong (2006B80407001) and the National Basic Research Program of China under Grant No. 2003CB314805

978-0-7695-3391-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICINIS.2008.18

networks (Section 3). Section 4 is devoted to an analysis of distance distributions in scale-free networks based on the results of Section 3. Section 5 contains our conclusions and some directions for further research.

2. Scale-Free Networks

A graph $G = (V, E)$ has a number of local and aggregate parameters that characterize its structure (regularity, modularity, dimension), connectivity (density, ease/speed of diffusion), and robustness (resilience to random or malicious faults). The parameters that are of interest in this paper include the following:

- M — Number of edges; $M = |E|$
- N — Number of vertices; $N = |V|$
- $d(v)$ — Degree of the vertex $v \in V$
- $\bar{d}$ — Average vertex degree of the network; $\bar{d} = \Sigma_{v\in V}\, d(v) / N$
- n_k — Number of degree-k vertices; $n_k = \left|\{v \mid d(v) = k\}\right|$
- $\delta(u, v)$ — Distance between vertices u and v
- D — Network diameter; $D = \max_{u,v\in V}\, \delta(u, v)$
- $\bar{\delta}$ — Average distance; $\bar{\delta} = \Sigma_{u,v\in V}\, \delta(u, v) / N^2$ [We use this more convenient definition, rather than excluding the case $u = v$ and dividing the sum of distances by $N(N-1)$]
- q_k — Number of node pairs that are of distance k; $q_k = \left|\{(u, v) \mid \delta(u, v) = k\}\right|$
- $P_d(k)$ — Degree distribution, or fraction of vertices that are of degree k; $P_d(k) = n_k / N$ [Probability that a randomly chosen vertex is of degree k]
- $P_\delta(k)$ — Distance distribution, or fraction of vertex pairs at distance k; $P_\delta(k) = q_k / N^2$ [Probability that the distance between a randomly chosen vertex pair is k]

For many complex networks, the probability distribution $P_d(k)$ of the number of degree-k vertices, also known as the degree distribution, can be represented (independent of scale, hence the designation "scale-free") by a power law with constant c and characteristic exponent γ :

$$P_d(k) = c\, k^{-\gamma} \qquad (1)$$

Equation (1) holds in an approximate manner, often providing greater precision for node degrees that are neither too small nor too large. Nevertheless, network attributes predicted when equation (1) is deemed to be exact often match remarkably well with the corresponding attributes of real complex networks. The characteristic exponent γ is in the range 2 to 3 for most actual complex networks.

Certain network attributes can be derived from equation (1). For example, the high resilience of scale-free networks in the face of random node failures and their vulnerability to deliberate attacks against selected nodes are consequences of equation (1), with certain mild assumptions. The fact that scale-free networks exhibit the small-world phenomenon, with the attendant high clustering coefficients, can be likewise deduced. Interestingly, the small-world phenomenon is present even if distances are not measured along shortest paths (which would require global knowledge for routing decisions) but rather according to a routing strategy with localized decisions [5]. Table 1 lists the parameters N, M, $\bar{d}$, and γ for several real scale-free networks [3, 4].

Table.1 Some parameters for several real scale-free networks

Network	N	M	$\bar{d}$	γ
Internet	10,687	31,992	5.98	2.5
Film actors	449,913	25,516,482	113.43	2.3
Metabolic network	765	3,686	9.64	2.2
Protein interactions	2,115	2,240	2.12	2.4

3. Necessary Conditions for Scale-Freedom

In this section, 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. Recall that $P_d(k)$ stands for the probability distribution of the number of degree-k vertices, $\bar{d}$ represents the average vertex degree, and n_k denotes the number of vertices of degree k. We have $M = ½ N\bar{d}$ and $n_k = N\, P_d(k)$, where N and M are numbers of vertices and edges, respectively. The preceding definitions imply:

$$\sum_{k=1}^{N-1} n_k = N \qquad (2a)$$

$$\sum_{k=1}^{N-1} k n_k = 2M \qquad (2b)$$

Supposing that $n_1 \neq 0$, a condition that is satisfied by many real complex networks, we have $n_k = n_1\, P_d(k) / P_d(1)$ and:

$$\sum_{k=1}^{N-1} [P_d(k) / P_d(1)] = N / n_1 \qquad (3a)$$

$$\sum_{k=1}^{N-1} [k\, P_d(k) / P_d(1)] = 2M / n_1 \qquad (3b)$$

For scale-free networks, we have $P_d(k) = P_d(1)\, k^{-\gamma}$, which leads to:

$$\sum_{k=1}^{N-1} k^{-\gamma} = N / n_1 \quad (4)$$

Therefore, using the assumption $\gamma \geq 2$, a condition which is known to hold for many scale-free networks [3, 4], we have:

$$N / n_1 \leq \sum_{k=1}^{\infty} k^{-\gamma} \leq \sum_{k=1}^{\infty} k^{-2} = \pi^2/6 \quad (5)$$

This leads to the result $N \approx n_1$. Let $f(k) = P_d(k) / P_d(1) = k^{-\gamma}$. When $N \approx n_1$, equations (3) yield:

$$\sum_{k=1}^{N-1} f(k) \approx 1 \qquad (6a)$$

$$\sum_{k=1}^{N-1} k f(k) \approx \overline{d} \qquad (6b)$$

Taking all logarithms to be in base 2, and in view of the assumption $\gamma \geq 2$, equations (6) yield $1 < \overline{d} \leq \log N$, 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 $\log N$.

Referring to Table 1, which lists the parameters N, M, $\overline{d}$, and γ for several real scale-free networks, we note that $\overline{d} \approx$ ½ log N (respectively, 5 log N, log N, or log log N) for the Internet (film actors, metabolic, or protein interaction) network.

For many real scale-free networks [3, 4], the exponent γ satisfies $\gamma < 3$. We now show that this too follows from our analysis. A connected scale-free network has $\gamma < 3$ when $n_1 \geq 1$ and $N \geq 6$. In fact, for $\gamma \geq 3$, we have:

$$M = (n_1/2) \sum_{k=1}^{N-1} k^{1-\gamma} \leq (\pi^2/12)\, n_1 < (5/6)\, n_1 \quad (9)$$

We see that the assumption $\gamma \geq 3$ leads to the impossible result $N - M > N/6 \geq 1$.

4. Distance Distribution

Using the method of Section 3, we can obtain other interesting results on distance distribution of complex networks. Recall that D and $\overline{\delta}$ represent the diameter and average distance of a complex network, respectively. Assume that there are q_k vertex pairs of distance k, where for simplicity we include the pairs (v, v) and count both (u, v) and (v, u) for $u \neq v$; that is, we consider the pairs to be ordered. Then we have:

$$\sum_{k=0}^{D} q_k = N^2 \qquad (10a)$$

$$\sum_{k=1}^{D} k q_k = N^2 \overline{\delta} \qquad (10b)$$

Clearly, $q_1 = 2M$. For scale-free networks with $\gamma \geq 2$, we have $M \leq$ ½ $N \log N$. These networks are also believed to satisfy $D \leq c \log N$, where c is constant. Now, suppose that $M =$ ½ $N \log N$, $D = \log N$, $\overline{\delta} =$ ½ $\log N$, and $g(k) = q_k/(2M)$. Then, equations (10) yield:

$$\sum_{k=0}^{\log N} g(k) \approx N / \log N \qquad (11a)$$

$$\sum_{k=1}^{\log N} k g(k) \approx N / 2 \qquad (11b)$$

The following identities involving binomial coefficients are well-known:

$$\binom{r}{0} + \binom{r}{1} + \binom{r}{2} + \cdots + \binom{r}{r} = 2^r \qquad (12a)$$

$$\binom{r}{1} + 2\binom{r}{2} + 3\binom{r}{3} + \cdots + r\binom{r}{r} = r\, 2^{r-1} \qquad (12b)$$

Based on equations (12), is easily seen that $g(k) = \binom{\log N}{k} / \log N$ satisfies equations (11). Hence, $q_k \approx N \binom{\log N}{k}$. Thus, we have obtained the distance distribution for vertex pairs in scale-free complex networks:

$$P_\delta(k) = q_k / N^2 \approx \binom{\log N}{k} / N \qquad (13)$$

In the case of certain regular networks, closed-form expressions for distance distribution can be obtained. For example, the corresponding formulas for $m \times m$ torus networks ($N = m^2$) are as follows:

$q_0 = N$; $q_k = q_{m-k} = 4kN$ $(1 \leq k < m/2)$ for all m (14a)

$q_{m/2} = (2m - 2)N$; $q_m = N$ for m even (14b)

Many real networks, with which we have experimented, have a distance distribution curve that rises at first, and then exhibits a declining trend, much like those in Fig. 1. So, both extremely short and very long distances are less common than values closer to the average. Figure 2 depicts the general form of the distance distribution in complex networks. Figure 3 depicts the distance distributions in two actual complex networks.

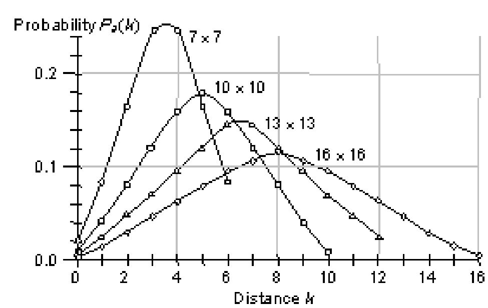


Fig. 1. Distance distribution for small, square torus networks.

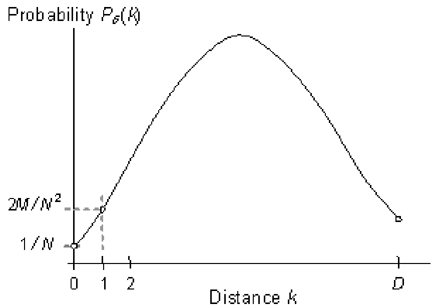
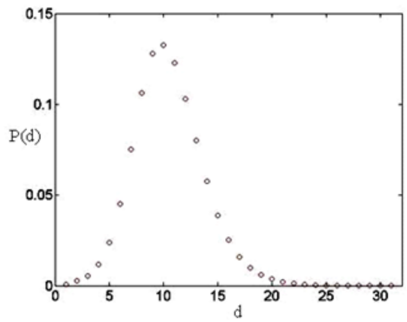
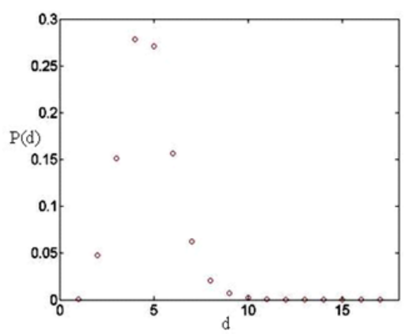


Fig. 2. Generic distance distribution in a diameter-D network with N vertices and M edges.



(a) Maximum component Δ of NCSTRL graph



(b) Maximum component Δ_1 of Linux graph

Fig. 3. Distance distribution in the maximum component of two actual complex networks: (a) NCSTRL graph [6] with 6396 vertices and diameter of 31, and (b) Linux graph [7] with 5285 vertices and diameter of 17.

5. 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 log N. Our method has the benefit of relying on conditions that are static and easily verified for any network. We have also derived a closed-form formula approximating the probability that a randomly chosen pair of vertices are at distance k (distance distribution) in scale-free networks. Such distributions are known to be applicable to diverse fields of study, including computer communication and software architecture. Our results on distance distribution are extensible to the case when distances are measured according to a routing strategy with localized decisions [5], in lieu of globally optimal routes which are much harder to compute.

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

References:

[1] A.-L. Barabási, and R. Albert, "Emergence of Scaling in Random Networks," *Science*, Vol. 286, pp. 509-512, 1999.

[2] D. J. Watts, and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, Vol. 393, pp. 440-442, 1998.

[3] R. Albert, and A.-L. Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, Vol. 74, pp. 47-97, 2002.

[4] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, Vol. 45, pp. 167-256, 2003.

[5] B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong, "Path Finding Strategies in Scale-Free Networks," *Physical Review E*, Vol. 65, Article 027103, 2002.

[6] M. E. J. Newman, "Scientific Collaboration Networks – I. Network Construction and Fundamental Results," *Physical Review E*, Vol. 64, Article 016131, 2001.

[7] C. R. Myers, "Software Systems as Complex Networks: Structure, Function, and Evolvability of Software Collaboration Graphs," *Physical Review E*, Vol. 68, Article 046116, 2003.