Average distance in a hierarchical scale-free network: an exact solution

Zhongzhi Zhang\textsuperscript{1,2}, Yuan Lin\textsuperscript{1,2}, Shuyang Gao\textsuperscript{1,2}, Shuigeng Zhou\textsuperscript{1,2} and Jihong Guan\textsuperscript{3}

\textsuperscript{1} School of Computer Science, Fudan University, Shanghai 200433, People’s Republic of China
\textsuperscript{2} Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, People’s Republic of China
\textsuperscript{3} Department of Computer Science and Technology, Tongji University, 4800 Cao’an Road, Shanghai 201804, People’s Republic of China

E-mail: zhangzz@fudan.edu.cn, sgzhou@fudan.edu.cn and jhguan@tongji.edu.cn

Received 17 September 2009
Accepted 14 October 2009
Published 28 October 2009

Abstract. Various real systems simultaneously exhibit scale-free and hierarchical structure. In this paper, we study analytically average distance in a deterministic scale-free network with hierarchical organization. Using a recursive method based on the network construction, we determine explicitly the average distance, obtaining an exact expression for it, which is confirmed by extensive numerical calculations. The rigorous solution obtained shows that the average distance grows logarithmically with the network order (number of nodes in the network). We show the similarity and dissimilarity in average distance between the network under consideration and some previously studied networks, including random networks and other deterministic networks. On the basis of the comparison, we argue that the logarithmic scaling of average distance with network order could be a generic feature of deterministic scale-free networks.

Keywords: exact results, random graphs, networks
1. Introduction

In the last decade, a lot of authors in different scientific communities have made a concerted effort toward unveiling and understanding the generic properties of complex networked systems in nature and society [1]–[4]. One of the most important discoveries is that despite network diversity, most real-life networks exhibit striking small-world behavior [5], that is to say, their average distance scales logarithmically with network order (number of nodes in a network), or slowly. Average distance is a fundamental measurement characterizing a complex network, which is relevant to many other structural features of the network, including degree distribution [6, 7], centrality [8], fractality [9]–[11], and so on. In addition, average distance has a strong effect on various dynamics running on networks, such as disease spreading [5], random walks [12], synchronization [13], amongst others. In view of its significance and usefulness, average distance has received considerable attention [14]–[20].

Apart from the small-world feature, a variety of real networks, particularly biological and social networks, also share two remarkable properties: scale-free behavior [21] and hierarchical structure [22, 23]. To mimic simultaneously the two prominent characteristics, Barabási, Ravasz and Vicsek proposed a deterministic model [24], hereafter called the BRV model, which is the progenitor of deterministic models for complex networks and has led to an increasing number of theoretical investigations on deterministic networks that are an interesting class of networks and have been proved to be a useful tool [25]–[36]. Many structural and dynamical properties of the BRV model have been studied in much detail, including the degree distribution [24], spectra of the adjacency matrix [37], random walks [38], to name but a few. However, in spite of its importance, the exact knowledge of the average distance for the BRV model remains not well understood.

To fill this gap, in this paper, we study the average distance in the BRV model, the deterministic nature of which makes it possible to investigate the average distance analytically. On the basis of recursive relations derived from the self-similar structure of the BRV model, we obtain the closed-form solution for the average distance. The rigorous result obtained shows that the average distance behaves logarithmically with the network order. This logarithmic scaling has also been previously reported for many other deterministic scale-free networks. We thus conjecture that the logarithmic scaling characterizes the behavior of the average distance for deterministic scale-free networks.

doi:10.1088/1742-5468/2009/10/P10022
2. The hierarchical scale-free network

Let us first introduce the BRV model, a hierarchical scale-free network that is constructed in an iterative way [24]. We denote by $H_g$ the BRV network model after $g$ ($g \geq 0$) iterations (number of generations). Initially ($g = 0$), the network $H_0$ consists of a single root node labeled as $i = 1$. At the first generation ($g = 1$), two new nodes $i = 2, 3$ are added to the system and connected to the root node. Thus, we get $H_1$, where node 1 is called the hub node, and the nodes $i = 2, 3$ are named bottom nodes, forming a set represented as $\mathbb{B}_1 = \{2, 3\}$. At generation 2 (i.e., $g = 2$), we generate two copies of $H_1$ and connect the bottom nodes of each replica to the hub of the original $H_1$. The hub of the original $H_1$ and the four bottom nodes in the replicas become the hub and bottom nodes of $H_2$, respectively. The set of bottom nodes belonging to $H_2$ is denoted as $\mathbb{B}_2$. Suppose one has $H_{g-1}$; the next generation network $H_g$ can be obtained from $H_{g-1}$ by adding two copies of $H_{g-1}$ with their bottom nodes being linked to the hub of the original $H_{g-1}$. In $H_g$, its hub is the hub of the original $H_{g-1}$, its bottom nodes are composed of all the bottom nodes of both copies of $H_{g-1}$, and all the bottom nodes make the set $\mathbb{B}_g$. Repeating indefinitely the replication and connection steps, we obtain the hierarchical scale-free network. Figure 1 illustrates the process of construction of the network for the first three iterations.

Some properties of the BRV model have been investigated in detail [37]. Let $N_g$ be the number of nodes in $H_g$, the network of the $g$th generation. By construction, at each new iteration, the number of network nodes increases by a factor of three, which together with the initial condition $N_0 = 1$ leads to $N_g = 3^g$. In $H_g$, the degree of all nodes and the number of nodes having the same degree can be determined exactly [37]. For example, the degree of the hub node is $K_h(g) = 2(2^g - 1)$, and the degree of the bottom nodes is $K_b(g) = g$. Again for instance, the cardinality, defined as the number of nodes in a set, of the set for the bottom nodes is $|\mathbb{B}_g| = 2^g$. The network is a sparse one with the mean degree averaged over all nodes being $\langle k \rangle_g = 4[1 - (\frac{2}{3})^g]$ which is approximately equal to 4 in the limit of infinite $g$. 

![Figure 1](image.png)
The BRV model presents some typical properties of real-life systems [24,37]. It is scale-free with the degree distribution exponent $\gamma = 1 + \ln 3/\ln 2$. In particular, the network has a crucial feature characterized by an obvious hierarchical structure that has also been observed in many real networks, e.g., metabolic networks [22,27]. All these characteristics are not shared by other previous models. The peculiar structural characteristics make the network unique within the category of scale-free networks. It is the precursor, probably the first model for hierarchical scale-free networks. However, in spite of its importance, rigorous knowledge of the average distance is still lacking; its exact determination is the primary topic of this paper.

3. Closed-form solution for the average distance

After introducing the hierarchical scale-free network, we now derive the average distance analytically. We represent all the shortest path lengths of network $H_g$ as a matrix in which the entry $d_{ij}(g)$ is the distance between nodes $i$ and $j$, that is the length of a shortest path joining $i$ and $j$. A measure of the typical separation between two nodes in $H_g$ is given by the average distance $d_g$ defined as the mean of distances over all pairs of nodes:

$$d_g = \frac{D_g}{N_g(N_g - 1)/2},$$

where

$$D_g = \sum_{i \in H_g, j \in H_g, i \neq j} d_{ij}(g)$$

denotes the sum of the distances between two nodes over all couples. Notice that in equation (2), for a pair of nodes $i$ and $j$ ($i \neq j$), we only count $d_{ij}(g)$ or $d_{ji}(g)$, not both.

We continue by exhibiting the procedure of determining the total distance and present the recurrence formula, which allows us to obtain $D_{g+1}$ for the $g + 1$ generation from $D_g$ for the $g$ generation. The hierarchical network $H_g$ has a self-similar structure that allows one to calculate $D_g$ analytically. According to the construction (see figure 2), network $H_{g+1}$ is obtained by joining three copies of $H_g$ that are labeled as $H^{(1)}_g$, $H^{(2)}_g$, and $H^{(3)}_g$. Using this self-similar property, the total distance $D_{g+1}$ satisfies the recursion relation

$$D_{g+1} = 3D_g + \Delta_g,$$

where $\Delta_g$ is the sum over all shortest path length whose endpoints are not in the same $H^{(\phi)}_g$ branch. The paths that contribute to $\Delta_g$ must all go through the hub node $X$, where the three copies of $H_g$ are connected. Hence, to determine $D_g$, all that is left is to calculate $\Delta_g$. The analytic expression for $\Delta_g$, referred to as the crossing path length, can be derived as below.

Let $\Delta_g^{(\alpha,\beta)}$ be the sum of the lengths of all shortest paths whose endpoints are in $H^{(\alpha)}_g$ and $H^{(\beta)}_g$, respectively. Then the total sum $\Delta_g$ is given by

$$\Delta_g = \Delta_g^{(1,2)} + \Delta_g^{(1,3)} + \Delta_g^{(2,3)}.$$  

By symmetry, $\Delta_g^{(1,2)} = \Delta_g^{(1,3)}$, so

$$\Delta_g = 2\Delta_g^{(1,2)} + \Delta_g^{(2,3)}.$$
Average distance in a hierarchical scale-free network

Figure 2. Schematic illustration of the means of construction of the hierarchical scale-free network. $H_{g+1}$ is obtained by joining three replicas of $H_g$ denoted as $H_g^{(\phi)} (\phi = 1, 2, 3)$, which are connected to one another at the hub node of $H_g^{(1)}$. The black node in the figure is the hub denoted by $X$ (not labeled).

Having $\Delta_g$ in terms of the quantities $\Delta_g^{(1,2)}$ and $\Delta_g^{(2,3)}$, the next step is to explicitly determine the two quantities.

To calculate the crossing distance $\Delta_g^{(1,2)}$ and $\Delta_g^{(2,3)}$, we give the following notation.

For an arbitrary node $v$ in network $H_g$, let $f_v(g)$ be the smallest value of the shortest path length from $v$ to any of the $2^g$ bottom nodes belonging to $B_g$, and the sum of $f_v(g)$ for all nodes in $H_g$ is denoted by $F_g$. Analogously, in $H_g$ let $h_v(g)$ denote the distance from a node $v$ to the hub node $X$, and let $M_g$ stand for the total distance between all nodes in $H_g$ and the hub node $X$ in $H_g$, including $X$ itself. By definition, $F_{g+1}$ can be given by the sum

$$F_{g+1} = \sum_{v \in H_g^{(1)}} f_v(g+1) + \sum_{v \in H_g^{(2)}} f_v(g+1) + \sum_{v \in H_g^{(3)}} f_v(g+1)$$

$$= \sum_{v \in H_g} [h_v(g) + 1] + 2 \sum_{v \in H_g} f_v(g)$$

$$= 2F_g + N_g + M_g, \quad (6)$$

and $M_{g+1}$ can be written recursively as

$$M_{g+1} = \sum_{v \in H_g^{(1)}} h_v(g+1) + \sum_{v \in H_g^{(2)}} h_v(g+1) + \sum_{v \in H_g^{(3)}} h_v(g+1)$$

$$= \sum_{v \in H_g} h_v(g) + 2 \sum_{v \in H_g} [f_v(g) + 1]$$

$$= M_g + 2(F_g + N_g). \quad (7)$$

Using $N_g = 3^g$, and considering $F_1 = 1$ and $M_1 = 2$, the simultaneous equations (6) and (7) can be solved inductively to obtain

$$F_g = 3^{g-2}(4g-1) \quad (8)$$

doi:10.1088/1742-5468/2009/10/P10022
and

\[ M_g = 2 \times 3^{g-2} (2g + 1). \]  

(9)

With above obtained results, we can determine \( \Delta_g^{(1,2)} \) and \( \Delta_g^{(2,3)} \), which can be expressed in terms of these explicitly determined quantities. By definition, \( \Delta_g^{(1,2)} \) is given by the sum

\[
\Delta_g^{(1,2)} = \sum_{u \in H_g^{(1)}, v \in H_g^{(2)}} d_{uv} (g + 1) \\
= \sum_{u \in H_g^{(1)}, v \in H_g^{(2)}} [h_u(g) + 1 + f_v(g)] \\
= \sum_{v \in H_g^{(2)}} \sum_{u \in H_g^{(1)}} h_u(g) + \sum_{u \in H_g^{(1)}} \sum_{v \in H_g^{(2)}} [1 + f_v(g)] \\
= N_g M_g + (N_g)^2 + N_g F_g.
\]

(10)

Inserting equations (8) and (9) into (10), we have

\[
\Delta_g^{(1,2)} = 9^g - 2(8g + 2).
\]

(11)

Proceeding similarly,

\[
\Delta_g^{(2,3)} = \sum_{u \in H_g^{(2)}, v \in H_g^{(3)}} d_{uv} (g + 1) \\
= 2[(N_g)^2 + N_g F_g] \\
= 9^{g-2}(8g + 8).
\]

(12)

Substituting equations (11) and (12) into (5), we get

\[
\Delta_g = 9^{g-2}(24g + 12).
\]

(13)

Substituting equation (13) into (3) and using the initial value \( D_1 = 4 \), we can obtain the exact expression for the total distance:

\[
D_g = 4g \times 9^{g-1}.
\]

(14)

Then the analytic expression for average distance can be obtained as

\[
d_g = \frac{8g \times 3^{g-2}}{3^g - 1}.
\]

(15)

We have checked our rigorous result provided by equation (15) against numerical calculations for different network orders up to \( g = 10 \) which corresponds to \( N_{10} = 59049 \). In all the cases we obtain a complete agreement between our theoretical formula and the results of the numerical investigation; see figure 3.

We continue to express the average distance \( d_g \) as a function of network order \( N_g \), in order to obtain the scaling between these two quantities. Recalling that \( N_g = 3^g \), we have \( g = \log_3 N_g \). Hence equation (15) can be rewritten as

\[
d_g = \frac{8N_g \ln N_g}{9 \ln 3 (N_g - 1)}.
\]

(16)
In the infinite network order limit, i.e., $N_g \rightarrow \infty$,

$$d_g = \frac{8}{9 \ln 3} \ln N_g.$$  \hfill (17)

Thus, for large networks, the average distance grows logarithmically with increasing order of the network.

This logarithmic scaling is similar to that of other hierarchical scale-free networks with high clustering coefficient, which was previously obtained in a quite different way by mapping the system onto a Potts model in one-dimensional lattices [28]. Since there is no triangle in the network studied, its clustering coefficient is zero. Our result, together with earlier work, shows that the clustering coefficient has no qualitative effect on the average distance of deterministic hierarchical scale-free networks, which is consistent with the phenomenon observed for random scale-free networks by using numerical simulations [39, 40].

However, the deterministic network under consideration also exhibits some different aspects as compared with the conventional (non-hierarchical) scale-free networks. For example, it has been suggested that for stochastic scale-free networks with degree distribution exponent $\gamma < 3$ and network order $N$, their average distance $d(N)$ behaves as a double-logarithmic scaling with $N$: $d(N) \sim \ln \ln N$ [6, 7], which is in sharp contrast to the logarithmic scaling obtained for the BRV model addressed here, in despite of the fact that the latter has a degree distribution exponent $\gamma = 1 + \ln 3/\ln 2$ less than 3. Actually, this logarithmic scaling of average distance with network order has also been shown in other deterministic scale-free networks with $\gamma < 3$ [19, 25, 28, 36, 41, 42]. Thus, deterministic scale-free networks present an obvious difference from their stochastic scale-free counterparts in the aspect of the structural property of the average distance. We speculate that the logarithmic scaling for the average distance can be used to establish the universality class for deterministic scale-free networks. Further studies are necessary.
to uncover the reasons for the similarity and dissimilarity of deterministic and random scale-free networks as regards average distance.

4. Conclusions

To conclude, scale-free behavior and hierarchical structure are ubiquitous in a variety of real-life systems. In this paper, we studied analytically the average distance of a deterministically growing scale-free hierarchical network introduced by Barabási, Ravasz and Vicsek [24], which can mimic some real-world networks to some extent. On the basis of the particular construction of the network, we obtained the rigorous solution for the average distance. We showed that in the infinite limit of network order $N_g$, the average distance $d_g$ exhibits a scaling law as $d_g \sim \ln N_g$. We also showed that there are similarities and dissimilarities of the behaviors of the average distances for deterministic and random scale-free networks. Finally, combining the result obtained and previous studies, we argued that the logarithmic scaling of average distance with network order may characterize deterministic scale-free networks.

Acknowledgments

This research was supported by the National Basic Research Program of China under grant No. 2007CB310806, the National Natural Science Foundation of China under Grants Nos 60704044, 60873040 and 60873070, Shanghai Leading Academic Discipline Project No. B114, and the Program for New Century Excellent Talents in University of China (NCET-06-0376).

References

[1] Albert R and Barabási A-L, 2002 Rev. Mod. Phys. 74 47
[2] Dorogovtsev S N and Mendes J F F, 2002 Adv. Phys. 51 1079
[3] Newman M E J, 2003 SIAM Rev. 45 167
[4] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwanga D-U, 2006 Phys. Rep. 424 175
[5] Watts D J and Strogatz H, 1998 Nature 393 440
[6] Chung F and Lu L, 2002 Proc. Natl Acad. Sci. USA 99 15879
[7] Cohen R and Havlin S, 2003 Phys. Rev. Lett. 90 058701
[8] Dorogovtsev S N, Mendes J F F and Oliveira J G, 2006 Phys. Rev. E 73 056122
[9] Song C, Havlin S and Makse H A, 2006 Nat. Phys. 2 275
[10] Zhang Z Z, Zhou S G and Zou T, 2007 Eur. Phys. J. B 56 259
[11] Zhang Z Z, Zhou S G, Chen L C and Guan J H, 2008 Eur. Phys. J. B 64 277
[12] Condamin S, Bénichou O, Tejedor V, Voituriez R and Klafter J, 2007 Nature 450 77
[13] Nishikawa T, Motter A E, Lai Y-C and Hoppensteadt F C, 2003 Phys. Rev. Lett. 91 014101
[14] Dorogovtsev S N, Mendes J F F and Samukhin A N, 2003 Nucl. Phys. 653 307
[15] Lovejoy W S and Loch C H, 2003 Soc. Netw. 25 333
[16] Fronczak A, Fronczak P and Holyst J A, 2004 Phys. Rev. E 70 056110
[17] Holyst J A, Sienkiewicz J, Fronczak A, Fronczak P and Suchecki K, 2005 Phys. Rev. E 72 026108
[18] Zhang Z Z, Zhou S G, Chen L C, Yin M and Guan J H, 2008 J. Phys. A: Math. Theor. 41 485102
[19] Zhang Z Z, Chen L C, Fang L J, Zhou S G, Zhang Y C and Guan J H, 2009 J. Stat. Mech. P02034
[20] Fekete A, Vattay G and Pósfai M, 2009 Phys. Rev. E 79 065101(R)
[21] Barabási A-L and Albert R, 1999 Science 286 509
[22] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási A-L, 2002 Science 297 1551
[23] Clauset A, Moore C and Newman M E J, 2008 Nature 451 98
[24] Barabási A-L, Ravasz E and Vicsek T, 2001 Physica A 299 559
[25] Dorogovtsev S N, Goltsev A V and Mendes J F F, 2002 Phys. Rev. E 65 066122
[26] Jung S, Kim S and Kahng B, 2002 Phys. Rev. E 65 056101
Average distance in a hierarchical scale-free network

[27] Ravasz E and Barabási A-L, 2003 Phys. Rev. E 67 026112
[28] Noh J D, 2003 Phys. Rev. E 67 045103(R)
[29] Noh J D and Rieger H, 2004 Phys. Rev. E 69 036111
[30] Hinczewski M and Berker A N, 2006 Phys. Rev. E 73 066126
[31] Rozenfeld H D, Havlin S and ben-Avraham D, 2007 New J. Phys. 9 175
[32] Hinczewski M, 2007 Phys. Rev. E 75 066104
[33] Zhang Z Z, Zhou S G, Zou T, Chen L C and Guan J H, 2007 Eur. Phys. J. B 60 259
[34] Zhang Z Z, Zhou S G, Fang L J, Guan J H and Zhang Y C, 2007 Europhys. Lett. 79 38007
[35] Comellas F and Miralles A, 2009 Physica A 388 2227
[36] Zhang Z Z, Guan J H, Ding B L, Chen L C and Zhou S G, 2009 New J. Phys. 11 083007
[37] Iguchi K and Yamada H, 2005 Phys. Rev. E 71 036144
[38] Agliari E and Burioni R, 2009 Phys. Rev. E 80 031125
[39] Holme P and Kim B J, 2002 Phys. Rev. E 65 026107
[40] Zhang Z Z, Roug L L, Wang B, Zhou S G and Guan J H, 2007 Physica A 380 639
[41] Zhang Z Z, Zhou S G and Chen L C, 2007 Eur. Phys. J. B 58 337
[42] Zhang Z Z, Chen L C, Zhou S G, Fang L J, Guan J H and Zou T, 2008 Phys. Rev. E 77 017102

doi:10.1088/1742-5468/2009/10/P10022