Self-similarity networks and self-similarity network group

Shaohua Tao*, Zhanshen Feng and Zhili Zhang

Department of Information Engineering, Xu Chang University, Xu Chang City, P. R. China

ABSTRACT

The self-similarity of complex has extensive practical background in real world. It is similar to the phenomena of social relationships: “things of one kind come together, birds of a feather flock together”. Hence, we proposed the self-similarity network evolving model based on attributes similarity between the nodes. In network each node has the attribute value, by this computing similarity between the nodes. If two nodes similar attribute falls in certain sector, then established the connection between nodes. The simulations make clear that the degree distribution of the self-similarity network similar to the small-world networks. The clustering and the average path of the self-similarity network are smaller than BA model and are bigger than small world. Similar network model is a new characteristic of complex network except the BA model and the small world mode. Similar network gathering similar network groups, and the information transfered quickly in similar network group.

Key words: self-similarity; attributes matching; information transfer; similar-degree; similar group

INTRODUCTION

In the last few years, complex networks study have become a focus of attention[1-7], which has extensive application on engineering technology, society and medicine[8]. Previous studies have primarily focused on finding various statistical properties of real networks, such as small world property[9,10,11], power law degree distribution[8], community structure[10,12,13] and hierarchical structure[14]. Many network models were proposed based on these characteristic, such as small world model[15], BA[11] and the BA extend model[16,17,18,19,20].While the small-world and scale-free network models capture the basic properties, it is the results that simplify the real world networks. It is necessary to further study the characters of complex networks which are more consistent with real world. In real world there has many network which nodes have some similar. For example, friend finder network makes friend as to objective, and commercial network that has some commercial nature. In making friends network, personal as nodes and the relationship between personal as edges. Between nodes each other has some common traits, namely the similarity, between the node will only have interactive. Hence, self-similarity of complex networks attracts the researchers interested gradually [21,22,23,24].

Chaoming Song advanced the self-similarity of complex networks, it used different edges of box covered the network, which obey the power law[25]. Although proposed the complex network has self-similarity, they haven’t revealed the microscopic evolved characteristic between the node and how to establish the connection. Moreover, we also raise the question that does similar network really obeys the power law distribution and has the BA nature? On further study we proposed the self-similarity concept of network is based on the similar characteristic between nodes which is forms by the node internal attribute microscopic interaction characteristic. Each node has certain attributes, which constituted set. Between nodes similar degree is a similarity, but the similarity decided by the node attribute collection. If two nodes attribute similarity falls in certain sector, then these two nodes have the similarity, established the connection and formed the similar characteristic of complex network. On the other hand, the node in transmit message's process, does not have the whole importance, but is effective to each sub-network. Each node belonging to different sub-networks, and various sub-network is similar or common nature of the nodes, these nodes
to link itself with a certain degree of similarity, therefore, sub-networks also have self-similar characteristics. We handle the set of sub-network as a group, therefore, self-similar network composed of sub-networks known as self-similar network group. Self-similar network group information transmitted will faster.

THE ATTRIBUTES SIMILARITY

First we make the simple introduction to the network attribute collection. The attribute and the attribute set in fuzzy theory have made the elaboration[26]. Here, we introduced the node attribute value and attribute set of the node in network detailed.

A. Preliminaries

Each node in networks has certain attributes. The attributes value of nodes can be described by two functions, which are true function and false function. The true and false functions associate a real number in the interval [0, 1]. All of the set of attributes make up of attributes set. Let \( v_p \) be a node, with a attribute \( e \), \( e \) is characterized by a true \( t_p \) and a false \( f_p \). \( t_p \) is a lower bound on the grade of \( e \) derived from the evidence for \( e \), and \( f_p \) is a lower bound on the negation of \( e \) derived from the evidence against \( e \). \( t_p \) and \( f_p \) both associate a real number in the interval [0,1] with each node attributes of \( v_p \), where \( t_p + f_p \leq 1 \).

B. The Single Attribute Matching of Nodes

The computation of node attribute value is as follows:

(1) Suppose \( e_p = [t_{e_p}, 1-f_{e_p}] \) is an attribute of node \( v_p \), \( t_{e_p} \in [0,1] \), \( f_{e_p} \in [0,1] \), and \( t_{e_p} + f_{e_p} \leq 1 \). then the value of \( e_p \) denoted by the function:

\[ C(e_p) = t_{e_p} - f_{e_p} \]
\[ C(e_p) \in [-1,1] \]

(2) Suppose \( e_p = [t_{e_p}, 1-f_{e_p}] \), \( e_q = [t_{e_q}, 1-f_{e_q}] \) is an attributes of node \( v_p, v_q \) respectively, the similar-degree of the pair of \( e_p, e_q \) can be calculated by the function \( S \):

\[ S(e_p, e_q) = 1 - \frac{(t_{e_p} - t_{e_q})^2 + (f_{e_p} - f_{e_q})^2}{2} \]

We can get the theorem as following:

Theorem 1: \( S(e_p, e_q) \in [0,1] \).

Theorem 2: \( S(e_p, e_q) = S(e_q, e_p) \)

C. Multi-attributes Matching of Nodes

The nodes \( v_p \) and \( v_q \) in network, the node \( v_p \) has the attributes with \( m \), and the node \( v_q \) also has the attributes with \( m \).

\[ v_p = (e_{p1}, e_{p2}, \ldots, e_{pm}) \]
\[ v_q = (e_{q1}, e_{q2}, \ldots, e_{qm}) \]

If the attributes \( e_p, e_q \) denoted as true and false, then the attributes set of the nodes \( v_p, v_q \) can shown as:

If the attribute element \( e_p, e_q \) can be denoted by true value and false value, the attributes set of the nodes \( v_p, v_q \)
denoted by

\[
v_p = \{t_p(e_{p_1}), 1-f(e_{p_1}), t_p(e_{p_2}), 1-f(e_{p_2}), \ldots, t_p(e_{p_m}), 1-f(e_{p_m})\}\]

\[
v_q = \{t_q(e_{q_1}), 1-f(e_{q_1}), t_q(e_{q_2}), 1-f(e_{q_2}), \ldots, t_q(e_{q_m}), 1-f(e_{q_m})\}\]

So, the similar-degree of nodes \(v_p\) and \(v_q\) can be calculated by the function \(T\).

\[
T(v_p, v_q) = \frac{1}{m} \sum_{i=1}^{m} S(v_p(e_{p_i}), v_q(e_{q_i})) = \frac{1}{m} \sum_{i=1}^{m} \left(1 - \frac{(t_{p_i} - t_{q_i})^2 + (f_{e_{p_i}} - f_{e_{q_i}})^2}{2}\right)
\]

Obviously, more bigger \(T(v_p, v_q)\) value, and more similar between the node \(v_p, v_q\).

## SELF-SIMILARITY NETWORK

### A. The Node Information Transfer

From above we know that the creating arithmetic of self-similarity network can be given according to the similar-degree between nodes. First each node can acquire its information and have some mutual effects among nodes and can pass information to each other. If the information is the same or similarity, then the nodes are connected and they are of one kind. Especially, more much similar between nodes, more easily have links. For example, people affiliate with somebody in the society, always find someone whether has the same with himself, such as character, favor and opinions about things. If someone has enough character with himself, then they will keep contact more usual. People always like affiliate with the one who has similarity with himself in all aspects, which is domino effect. It can be captured as follows in networks:

1) The node \(v_p = (e_{p_1}, e_{p_2}, \ldots, e_{p_m})\) perceives and gets the information of it. Supposed it has \(m\) features \((e_{p_1}, e_{p_2}, \ldots, e_{p_m})\), and node \(v_p\) pass the information to the nodes which around it. If the information passed by nodes has the similarity, then the nodes are connected.

2) The new nodes \(v_q = (e_{q_1}, e_{q_2}, \ldots, e_{q_m})\) occur to networks every \(h\) time, the new nodes and old nodes pass the information to another, if the new nodes have the similarity information with any old nodes in networks then connection is created.

3) If the attributes \((e_{q_1}, e_{q_2}, \ldots, e_{q_m})\) of node \(v_q\) and attributes \((e_{p_1}, e_{p_2}, \ldots, e_{p_m})\) of \(v_p\) have the similarity information then denotes as: \(\forall i \in \{1, 2, \ldots, m\}\), \(e_{q_i} \sqcap e_{p_i}\), and the \(e_{p_i} = [t_{p_i}, 1-f_{e_{p_i}}], e_{q_i} = [t_{q_i}, 1-f_{e_{q_i}}]\).

4) The probability that node \(v_p\) connects with \(v_q\) depends on similar degrees, the linking probability is following:

\[
\Pi(v_p, v_q) = \frac{1}{m} \sum_{i=1}^{m} \left(1 - \frac{(t_{q_i} - t_{p_i})^2 + (f_{e_{q_i}} - f_{e_{p_i}})^2}{2}\right)
\]

To simple, the connection probability equation (1) expressed by \(p\).

From above we can known, more close to 1 the \(T(v_p, v_q)\) value, more similar the nodes \(v_p, v_q\). Hence, we assume that the connection probability \(p \in [0.7, 1]\), the new nodes and the old are connected.

188
B. The Node Clustering Transfer Information

The old nodes link the new nodes according to equation (1), which formed the node clustering \( V_1, V_2, \ldots, V_i \), \( i \in \{1, 2, \ldots, N\} \), then \( V_1 = \{v_1, v_2, \ldots, v_n\}, V_2 = \{v_1, v_2, \ldots, v_n\}, \ldots, V_i = \{v_1, v_2, \ldots, v_n\} \). The node clustering also can transfer information each other, the node clustering \( V_j, i \in \{1, 2, \ldots, N\} \) transfer the information to the vicinity nodes clustering \( V_j, j \in \{1, 2, \ldots, N\} \), if both of attributes value are similar then has a link. Multi-node clustering connection constitutes the self-similarity of local network. If the node added continually, and the node clustering also added continually which formed one or multi-local network, and finally formed the network similarity, shown as Fig.1.

![Node Clustering and Local Self-Similarity Network](image)

Fig.1 The node clustering and the local self-similar network

C. Network Evolving

We used the pajek tool to draw the different similar network, the node number and the connection probability respectively are \( N = 200, p = 0.468 \), \( N = 500, p = 0.8 \), \( N = 500, p = 0.768 \), \( N = 1000, p = 0.8 \), shown as Fig.2.

![Network Evolving](image)

(a) \( n=200 \)  \( p=0.468 \)

(b) \( n=500 \)  \( p=0.8 \)
Obviously from Figure 2, when the connects similarly probability is small, the network will have many edge connects. Considered real network similar connection, we have established the valve value to the similar connection probability. When $S(e_p, e_q) \geq 0.85$, two nodes has edge, therefore, we simulated similar connection probability, shown as Fig.4. Between the node connection probability is located between $[0.85,1]$.

**SIMULATION AND ANALYSIS**

A. Measures of Self-similarity

Network is dynamically growing in real world. To measure self-similarity of scale-free networks, take the growing network as an example to calculate $D_r$ at different stages\(^\text{[26]}\). The networks at steps are: initialization of local networks, growth of local networks and finally generated networks.

**Step 1:** Cover the network with squares whose edge length is $r$, calculate how many sub networks are contained in self-similarity network groups, then register them in $N_r(F)$.

**Step 2:** Decrease edge length and again calculate how many sub networks are contained in self-similarity network groups, then register them in $N_r(r)$ and so on.

**Step 3:** Calculate the registered $N_r(r)$ of various $r$.

Volume-dimension $D_{r1}$, $D_{r2}$, $D_{r3}$ and $D_{r4}$ were calculated respectively at different stages. The network has self-similarity property if $D_{r1}$, $D_{r2}$, $D_{r3}$ and $D_{r4}$ have the same or nearly the same values.

As some examples, we use the networks with number of nodes of $n = 100$, $n = 500$, $n = 1000$, $n = 1500$, respectively. The network of $n = 100$ is a small local network, then add nodes to $n = 500$ and $n = 1000$, we get large local networks. The networks continually grow until $n = 1500$. The box counting is used to measure $N_r(r)$ of various $r$ in local networks. The simulation result is shown as Fig.3.
Linear fit the generated data to $D_c = \lim_{r \to 0} \left( \frac{\log N(r)}{\log \left( \frac{1}{r} \right)} \right)$ using the least square method and the resulting slope of the function line is the volume-dimension $D_c$. $D_{c1} = 1.56$ when $n = 100$; $D_{c2} = 2.98$ when $n = 500$; $D_{c3} = 2.85$; $D_{c4} = 2.85$ when $n = 1000$; and $D_{c4} = 2.41$ when $n = 1500$. The serial values of
$D_{c2}$, $D_{c3}$ and $D_{c4}$ reveal that the growing complex networks indeed have self-similarity in real world. However, there exists a deviation that $D_{c1} = 1.56$ when $n = 100$. This even reveals that self-similarity is more obvious during growth of the networks. Moreover, it can be shown that the power-law distribution (see Fig.3) is in accordance with the property of scale-free networks.

Obviously, power-law is not always the results of preferential connection. The simulation and analysis is processed under the network growth continual, which is also accordance with the real networks that continually growth.

### B. Degree Distribution

According to above, we assume the connection probability $p \in (0.7, 1)$, namely when the connection probability is bigger than 0.7, then two nodes only have the enough similarity and establish the connection. To simple, here the SN represents the self-similar network. The topology of the SN is simulated by Pajek and the degree distribution is calculated. From figure 2 can seen that the degree distribution of the SN network is similar to the small world. The curve has a peak value, and the value nearby deviation peak is very quickly weakened. However, the degree distribution of the BA is the power-law, shown as Fig.4.

![Image](https://example.com/figure4.png)

**FIG.4:** The degree distribution of the SN network and the BA network. The total number of node is 5000. The horizontal is the number of node, and the vertical is $p(k)$. Figure (a) shows the degree distribution of the SN, and the curves show the degree distribution at $p = 0.7$, $p = 0.75$, $p = 0.768$, $p = 0.8$, $p = 0.825$, separately. Figure (b) shows the degree distribution of the BA. The initial values are $m_0 = 0.1$, $m_0 = 0.3$, $m_0 = 0.5$, $m_0 = 0.7$, separately.

### C. The Clustering and Average Path

We compared the clustering and the average path length of the SN with the BA network. From the figure 3 we can known that the clustering of the SN is between $[0.047, 0.0505]$, and to the same size of network the clustering of the BA is between $[0.01, 0.03]$. So the clustering of the SN is higher than the BA. To the same size of the network the average path length of the SN is between $[2.798, 2.044]$, and the average path length of the BA is between $[2.998, 3.765]$. We can see that the average path length of the SN is shorter than the BA. From this may obtain the network which forms based on certain similarity to be easier to gather, therefore, the average path length of the SN network is shorter than the BA network, shown as Fig.5.
D. The Robust
The network behavior and nature has the closed relation with the network structure. If the network was removed the few nodes, and the major node was still connection, then we said that the network has the robustness against the node breakdown. The paper considers the two kind of node removing method. First, the random fault, namely removes the node completely random. Second, attacks deliberately, namely consciously removes the node that has the major link with the other nodes. Suppose the proportion which the number of node was removed compared to the total number of node in network was \( f \). We can measure the network robustness by using the size of the biggest connected sub graph \( s \) and using the relations of the average path length \( l \) and \( f \). The research discovered that the SN network and BA have the obviously difference to random and deliberately breakdown. Regarding the random fault, the SN network and the BA may maintain the basic connectivity. But the SN network to random fault has the good robustness compared to the BA network. The biggest connection sub graph of the SN network and the BA drops zero when \( f \) is relatively high; and the growth of the average path length of the SN is lower than the BA, figure 3. To deliberate attack, the SN network has the robustness similarly. Attack the node that has the major link with other node, the biggest connected sub graph's relative size still maintained certain proportion, the network still maintains the normal operation. But the BA network to attacks deliberately is quite frail. Attacked the nodes which have the major links in network, then the BA network may divided into many isolate sub graph, Fig. 6(b) is removed the few nodes have the highest degree in SN and BA networks. This is mainly because between the SN node is the similar connection, no has hubs nodes in network. To deliberately attack, the average path length growth of the SN is similar slower than the BA network. The paper proposed the self-similar model which is based on the similar values link between node, compared to the preference connection of the BA which enhanced the robustness to random fault and malice attack.
FIG. 6: the robustness and frail of the SN and BA. The horizontal is $f$, the vertical is $S$. Figure (a) shows the relationship between $S$ and $f$ to random attack. Red ($\square$) is the SN network, and the blue ($\circ$) is the BA network. Figure (b) shows the relationship between $S$ and $f$ of the SN and the BA to deliberately attack. Red ($\square$) shows the SN network, and the black ($\blacklozenge$) shows the BA network. Figure (c) shows the relationship between $I$ and $f$ to random attack. The horizontal is $f$, the vertical is $I$. Red ($\blacklozenge$) is the SN network, and the blue ($\blacklozenge$) is the BA network. Figure (d) shows the relationship between $I$ and $f$ to deliberately attack. Red ($\blacklozenge$ is the SN network, and the blue ($\circ$) is the BA network.
PROPERTIES OF SIMILARITY NETWORKS GROUP

A. The properties of Local Network

$N_1, N_2, ..., N_i$ are the subgroup of similar network group, thus we discussed $G$ changed to the $N_1, N_2, ..., N_i$, therefore transforms the child network into the corresponding same step Toeplitz square formation.

The nature 1 matrix has the interchangeability, namely to random $N_i \in G$ has

$N_iN_2 = N_2N_i$

For example,

$N_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 2 & 1 & 1 \end{bmatrix}, \quad N_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 0 & 2 & 1 \end{bmatrix}$

$N_iN_2 = N_2N_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 7 & 2 & 1 & 0 \\ 6 & 4 & 3 & 1 \end{bmatrix}$

The nature 2 matrix existence unit element, namely to any $N_i \in G$, $N_iE = EN = N_i$

For example,

$N_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 2 & 1 & 1 \end{bmatrix}$

$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

$N_iE = EN_i = N_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 2 & 1 & 1 \end{bmatrix}$

The nature 3 elements in $N_i$ are the inverse element, namely to any $N_i \in G$, exists $N_i^{-1} \in G$.

$N_i^{-1}N_i = N_i^{-1}N_i^{-1} = E$

And, its inverse matrix was still the Toeplitz matrix.
B. Union Set and Intersection of Similar Network Group

The network groups have the sub-network \( N_1, N_2, \ldots, N_i \), suppose the set of sub-network \( N_i \) is the closed subset of \( R \), mapping \( S : N_i \rightarrow N_i \) called the compression on the network. If to all \( v_i, v_j \) on \( N \), has a number \( c \), satisfied \( 0 < c < 1 \), then has the following theorem:

Theorem 1: suppose \( N_1, \ldots, N_n \) is compress mapping on \( N_i \subseteq R^n \), have

\[
|S_i(v_i) - S_i(v_j)| \leq c |v_i - v_j|, \quad \{v_i, v_j\} \in N_i,
\]

, to each \( i, c_i < 1 \), then only exists invariable compact subset \( S \) to the \( N_i \), satisfied

\[
S = \bigcup_{i=1}^n N_i
\]

From theorem 4 may see the whole network is constituted by union of set of the sub-network, shown as Fig.7.

![Fig.7 Union set](image)

These union of set do not have the intersection, \( N_i \) is surely all does not connect. But in the reality network has the connectivity, therefore, the set of sub-network will also satisfy the following condition:

**Theorem 2** suppose \( N_1, \ldots, N_n \) is compact mapping on \( N_i \subseteq R^n \),

\[
|S_i(v_i) - S_i(v_j)| \leq c_i |v_i - v_j|, \quad \{v_i, v_j\} \in N_i,
\]

to each \( i, c_i < 1 \), existing intersection set \( M \), satisfy:

\[
M = \bigcup_{i=1, j=1}^n \left( N_i \cap N_j \right), i \neq j
\]

Namely has non-empty intersection of set

\[
M_i = N_i \cap \bigcap_{j=1}^n N_j
\]

Existence like this some set of network,

\[
M_1 = N_1 \cap N_2 \neq 0, \quad M_2 = N_1 \cap N_3 \neq 0, \quad \ldots, \quad M_{i-1} = N_i \cap N_{i-1} \neq 0, \quad \ldots,
\]

Certainly also has the possibility to have intersection of set is \( 0 \), causes \( M_3 = N_2 \cap N_3 = 0, \quad M_6 = N_6 \cap N_7 = 0, \quad \ldots, \quad M_{i+1} = N_1 \cap N_{i+1} = 0 \), shown as Fig.8.

\( N_1, N_2, N_3 \) intersection with \( N_4 \) separately, then forms the node set of intersection network \( M_1, M_2, M_3 \).

![Fig.8 Intersection set](image)
The network has the similarity will cause in the network the information transmission to be quicker, we will use the network the average shortest path to indicate the information transfer.

CONCLUSION

The attributes value, single property matching and attributes matching between nodes in networks are described in detail; and more big attributes value of nodes, more similar of nodes. The node could transfer information, if the similar-degree values fall into certain sector, the two nodes linked each other. The self-similarity network model based on information transfer is given and the creating arithmetic of this model is also captured. The self-similarity network model detail; and more big attributes value of nodes, more similar of nodes. The node could transfer information, if the

Acknowledgments

This work was supported in part by a grant from natural foundation of Henan(No.122102210544 , No.12A510022 and No. 132300410349 ) and natural foundation (No. U1304403).

REFERENCES

[23] Chaoming Song, Shlomo Havlin and Hern_an A. Makse, Complex networks are self-similar, nature 433, 2004.