Communication Network Role Analysis Based on Weighted Directed Topological Potential

Wan-Yu Li¹, a, Guang-Quan CHENG¹, *, Jin-Cai HUANG

¹Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha, China

*18900728027@163.com

*Corresponding author

Keywords: Complex network, Role analysis, Communication network, Topological potential.

Abstract. Role analysis is a very important branch of complex network research, which tries to identify the node types with similar structures or functions. Inspired by the midfield in physics, and by extending and further expanding the concept of topological potential, a role identification and analysis method based on weighted directed topological potentials in connection with node “mass” (i.e. node own attribute) was introduced in this paper. This proposed method takes into consideration both the node own attributes and the network topological attributes, and identifies node roles by the difference in node topological potential. Its effectiveness was well verified by the experimental results on an IP communication network.

Overview

With the rapid development of the theoretical research of complex networks and the real complex networks, research focus has been shifted from initial network models to micro domains, and breakthroughs have been achieved in community mining, node influence and importance sequencing et al. But little efforts have been made on the analysis of node roles. The so-called role of node refers to the position or function of a node in a network. The role analysis is to divide the nodes of a whole network into types by similar structure or similar function, and then interpret the roles of various types.

Scott [1], an authoritative scholar in network analysis field appointed out long ago: the role analysis of nodes is a core component in network analysis. The accurate identification and analysis of the node roles in network are of great practical significance. For example, for a criminal network, it is hoped to more accurately identify all suspects, track and capture key members so as to effectively destroy the whole criminal network [2,3]. For social networks, it is hoped to analyze the role characteristics and behaviors of individuals so as to better plan the marketing, control the public opinions, and optimize the network services [4].

At present, the research efforts on the node roles of complex network are limited, and mainly focus on two aspects: network structure-based analysis and node attribute-based analysis. In the former case, research work generally starts from the structural attributes and similarities of network. Some researchers utilize node centrality indicators (degree centrality, between centrality, closeness centrality, subgraph centrality [5], k-shell centrality [6], collective influence [7]) for node analysis. Some researchers conduct analysis based on structural equivalency, and the similarities between nodes were measured in light of various structural similarity indicators. Lv Linyuan [8] has made a complete summary of similarity indicators, such as Jaccad index, cosine similarity and Salton indicator.

In the node attribute-based analysis of node roles, the internal and external attributes of a node are quantified into its characteristic indicators, and various characteristics of node are then analyzed. The research efforts on node attribute-based role analysis are in varied forms, and each has its merits. For example, Zhu Tian [9] clustered the user social characteristics obtained from social network analysis, and divided users into roles of general users, potential users and authoritative users.
It is hard to say that a network in real application represents only network structural characteristics or node attribute characteristics, and therefore, in node role identification and analysis, both of the two kinds of characteristics should be taken into consideration so as to improve the efficiency and accuracy of role identification. For this end, taking IP communication network as an example, this paper takes into consideration both network structural characteristics and node attributes, and constructs the weighted directed topological potential of node as an indicator for node role classification, and puts forward a node role identification method based on weighted directed topological potential.

**IP Communication Network Role Analysis Based on Weighted Directed Topological Potential**

**Undirected Topological Potential and Weighted Directed Topological Potential**

The concept of field was first put forward in physics to describe the interaction between matter particles. By reference to the concept of midfield in physics, the description method for interaction between matter particles, and their field and potentials are introduced into the abstract number field space. In an interaction field, each object in the field will be influenced by other objects, and thus, this space constitutes a data field. GAN WY [10] applied this data field to an undirected network graph \( G \) including \( n \) nodes and \( m \) edges, where the edge represents there is connection between two points, analogy to field potential, any point in \( G \) will be affected by other points in the field.

Network \( G = (V, E) \), where, \( V \) is a node set, \( |V| = m \), \( E \) is an edge set, \( |E| = m \), according to the definition of potential function of data field, the topological potential of node \( v_i \) may be expressed as Eq.1:

\[
\varphi(v_i) = \sum_{j=1}^{n} m_j \times e^{-\frac{d_{ij}\sigma}{\sigma}}
\]

(1)

Where, \( d_{ij} \) represents the network distance between node \( v_i \) and node \( v_j \), which is measured by the shortest route length; \( m_j \geq 0 \) represents the mass of node \( v_j \), and may be used to characterize the intrinsic attribute of each node.

\( \sigma \) means the impact factor, which is used to control the scope of influence of each node. Here, refering to GAN WY’s study [11], the potential entropy of topological potential field may be used to optimize \( \sigma \). For the optimization of \( \sigma \) value, the corresponding \( \sigma \) at the least entropy \( H \) should be selected. The corresponding potential field can be defined as Eq.2.

\[
H = -\sum_{i=1}^{n} \frac{\varphi(v_i)}{Z} \log\left(\frac{\varphi(v_i)}{Z}\right)
\]

(2)

Where, \( Z = \sum_{i=1}^{n} \varphi(v_i) \) is a standard factor.

Topological potential function is in essence a Gaussian function, and in light of the property of Gaussian function, it can be known from “3\( \sigma \)” rule that the functional scope of each node is the 3\( \sigma/\sqrt{2} \) adjacent space with this node as the center. When \( \sigma \in (0, \frac{\sqrt{2}}{3}) \), the interaction between nodes is 0, the potential value of each node is \( 1/n \), and the potential entropy takes the maximum log (n); when \( \sigma \in \left[\frac{2\sqrt{2}n}{3}, \frac{2\sqrt{2}(n+1)}{3}\right] \), a node affects its neighbor nodes within n-hop distance.
When $\sigma \in \left[ \frac{L}{\sqrt{2}}, +\infty \right)$, $L$ is the network diameter, and there is interaction even between the node pair with the longest space. Therefore, the interaction among all nodes should be considered in calculating the topological potential of any node.

The above discussion is based on a non-weighted and undirected network diagram. However, for any real network, the directionality of node connection and edge weight value are mostly considered, and these attributes may influence the potential value of a node. Therefore, for expanding an undirected topological potential, the weighted directed topological potential of node is built. This potential defines the outgoing potential and incoming potential of each node, the out-degree topological potential of a node mainly embodies the impact of this node on other nodes while the in-degree topological potential of a node mainly exhibits the impact of other nodes on this node. Besides, we add the influence of edge’s weight to the measurement of distance between nodes, and the weight value influences the strength of interaction between nodes. Therefore, edge weight value is introduced into the calculation of the shortest route between two nodes, $wd_{ij}$ is used to represent the weighted shortest route between node $i$ and $j$, and the in-degree and out-degree topological potential is calculated as follows:

\[
\begin{align*}
\varphi_{out}(v_i) &= \sum_{j=1}^{n} m_j \times e^{-\frac{wd_{ij}}{\sigma}} \\
\varphi_{in}(v_i) &= \sum_{j=1}^{n} m_j \times e^{-\frac{wd_{ji}}{\sigma}}
\end{align*}
\]

$m_j$ represents the mass of node $j$, an attribute indicator calculated based on node attribute.

For an actual situation, the larger the weight value is, the greater the interaction between two nodes will be, which means that the interaction is positively correlated with the edge weight value; the shorter the distance between two nodes, the greater the interaction will be, which means that the interaction is negatively correlated with the space between two nodes. Then, $wd_{ij}$ is used to represent the weighted shortest route value between $v_i$ and $v_j$:

\[
wd_{ij} = \frac{d_{ij}}{w_{ij}}
\]

Where, $d_{ij}$ represents the length of the shortest route between $v_i$ and $v_j$. Assume that the length of each edge is 1, $d_{ij}$ will be the number of edges $d$ included in the shortest route. $w_{ij}$ represents the sum of the edge weighted values of edges that the shortest route passes. If the shortest route between $v_i$ and $v_j$ passes edge $e_k$, $w_k$ represents the weighted value of edge $e_k$, then

\[
\frac{d_{ij}}{w_{ij}} = \frac{d}{\sum_{i} w_k}
\]

Role Identification Procedure of IP Communication Network

The network communication data of several network nodes in a certain period as shown below were processed to obtain the data structure as shown in Tab.1 (partially displayed):
Table 1. Partial Data of IP Communication Network.

<table>
<thead>
<tr>
<th>SRCIP</th>
<th>DSTIP</th>
<th>SRC PORT</th>
<th>DST PORT</th>
<th>FILE LEN</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.24.248.253</td>
<td>10.54.240.5</td>
<td>445</td>
<td>3822</td>
<td>1765</td>
<td>9/12 2:45:44</td>
</tr>
<tr>
<td>10.18.112.246</td>
<td>10.65.216.106</td>
<td>445</td>
<td>4364</td>
<td>2628</td>
<td>9/12 2:45:44</td>
</tr>
<tr>
<td>10.54.216.221</td>
<td>10.65.216.106</td>
<td>445</td>
<td>4365</td>
<td>270</td>
<td>9/12 2:45:45</td>
</tr>
</tbody>
</table>

SRCIP refers to source IP; SRCPORT refers to a port of source IP; DSTIP refers to destination IP; DSTPORT refers to a port of destination IP; FILELEN refers to the data size transmitted from the source to the destination. Each data is transmitted from a certain port of source IP to a certain port of destination IP. This data set has altogether 4999 records, involving 157 IPs and 1898 ports.

The role identification method adopted in this paper classifies node roles by the similarity of node topological potential. The specific procedure are as follows:

1. Construct a calculation formula for the directed topological potential of IP node with a node mass of \( m \) and an edge weight of shortest distance \( w_{ij} \).

   \[
   m_i = \frac{P_i}{P} \times \frac{\text{len}_i}{\text{LEN}} \tag{7}
   \]

   \( p_i \) represents the total number of intrinsic ports of node \( v_i \), \( P \) represents the total number of all IP ports, \( \text{len}_i \) is the sum of in-and-out traffic of node \( v_i \), and \( \text{LEN} \) is the total data traffic of this communication network in a given period. So the directed topological potential of IP node can be calculated.

2. Discuss the value taking for parameter \( \sigma \): determine the \( \sigma \) value that minimizes \( H \) according to chapter 2.1. This \( H \) is the sum of in-degree topological potential entropy and out-degree topological potential entropy.

3. Calculate the weighted directed topological potential of each node: \( \phi_{\text{out}}(v_i) \) and \( \phi_{\text{in}}(v_i) \).

4. Role classification: define the typical roles in the area with \( \phi_{\text{out}}(v_i) \in [\phi_{\text{out}}(v_i)_{\text{min}}, \phi_{\text{out}}(v_i)_{\text{max}}] \) and \( \phi_{\text{in}}(v_i) \in [\phi_{\text{in}}(v_i)_{\text{min}}, \phi_{\text{in}}(v_i)_{\text{max}}] \):
   
   \[ \text{A} \ (\phi_{\text{out}}(v_i)_{\text{min}}, \phi_{\text{in}}(v_i)_{\text{min}}), \text{B} \ (\phi_{\text{out}}(v_i)_{\text{max}}, \phi_{\text{in}}(v_i)_{\text{min}}), \]
   
   \[ \text{C} \ (\phi_{\text{out}}(v_i)_{\text{max}}, \phi_{\text{in}}(v_i)_{\text{max}}), \text{D} \ (\phi_{\text{out}}(v_i)_{\text{min}}, \phi_{\text{in}}(v_i)_{\text{max}}). \]

5. Calculate the Euclidean distance of each node to the four typical nodes, and then, assign each node to the typical role with the least distance.

Experiment

According to the above formulas, the in-degree and out-degree topological potential are calculated with the optimal \( \sigma \) value 0.975.

Fig.1 shows the topological potential, four typical roles (A, B, C, D) and the final recognition of roles. In this double logarithmic coordinate, the x-axis is the out-degree topological potential of each node, and the y-axis is the in-degree topological potential of each node.

Fig.2 shows all the nodes, all the communication between nodes and roles division of a network. Each nodes stands for each IP and each directed edges stands for the communication between the IP. Edge’s thickness indicates how many times they communicate with each other, while the more times they communicate, the thicker the edge is. The color of different roles is consistent with Fig.1.
Then four roles are represented by four colors:

1. The general node: the red nodes in the figure. In-of this type of nodes are both relatively small. They are not much active nodes in the network.

2. The main receiving node: the black nodes in the figure. The in-degree topological potential of this role is relatively large while out-degree topological potential is relatively small. These nodes receive large amount of information in the network and are greatly influenced by the rest of nodes in the network.

3. The main sending node: the green nodes in the figure. Compared to the other nodes, the out-degree topological potential of this role is relatively large while in-degree topological potential is relatively small. These nodes send large amount of information to other nodes and greatly influence the rest of nodes in the network.

4. The main transformation node: the blue nodes in the figure. The in-degree and out-degree topological potential of this type of nodes are both relatively big, on the one hand, they receive information, on the other hand, they send messages.

Summary

This paper highlights the role analysis of communication network nodes, and has solved the problems in the role identification and network nodes analysis. The paper uses “mass” to describe node characteristics and weighted directed topological potential to characterize the interaction among nodes, classifies the node roles by the similarity of nodes topological potential, and fully integrates the attributes of network nodes and those of network structure. In further role division based on the obtained weighted directed topological potential, the number of classified roles is limited, and a more general method may be used in the future research so as to achieve better and more general effects.

Reference


