

Communication Network Dynamic Role Analysis Based on Weighted Directed Topological Potential

Wan-Yu Li^{1,a}, Guang-Quan CHENG¹, Jin-Cai HUANG¹

¹Science and Technology on Information Systems Engineering Laboratory,
National University of Defense Technology, ChangSha, 410073, China

^aemail:18900728027@163.com

Keywords: complex network, topological potential, communication network, role analysis

Abstract. Role analysis is a very important branch of complex network research, which tries to identify the node types with similar structures or functions. By extending 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. Broke through the limitations of static network role analysis and proposed a new method based on the change of nodes in time series. Its effectiveness was well verified by the experimental results on an IP communication network.

Introduction

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.

Role analysis can be applied to the study of crime networks, social networks, communication networks and other fields^[1-4]. The method of the role analysis may include two aspects: One is the network structure-based analysis, which is based on the network structure equivalence and the nodes similarity index to identify the nodes with similar structure^[5-8]. The other one is the node attribute-based analysis. the internal and external attributes of a node are quantified into its characteristic indicators, and various characteristics of nodes are then analyzed^[9,10]. Most of the traditional research on the role analysis aimed at static network, or took a network as a static network for a period of time and then fixed the node's property together. The role identification of the dynamic characteristics of nodes in the network^[11,12] is not much, however, the role character of a node should be the stable and regular characteristic during a long time. So we need to analyze the role of nodes in the dynamic process of the network.

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. GAN WY^[13,14] 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:

$$\varphi(v_i) = \sum_{j=1}^n (m_j \times e^{-\left(\frac{d_{ij}}{\sigma}\right)^2}) \quad (1)$$

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.

σ means the impact factor, which is used to control the scope of influence of each node. Here, referring to GAN WY's study, the potential entropy of topological potential field may be used to optimize σ . For the optimization of σ value, the corresponding σ at the least entropy H should be selected. The corresponding potential field can be defined as Eq.2.

$$H = -\sum_{i=1}^n \frac{\varphi(v_i)}{Z} \log\left(\frac{\varphi(v_i)}{Z}\right) \quad (2)$$

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

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. 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:

$$\varphi_{out}(v_i) = \sum_{j=1}^n m_j \times e^{-\left(\frac{wd_{ij}}{\sigma}\right)^2} \quad (3)$$

$$\varphi_{in}(v_i) = \sum_{j=1}^n m_j \times e^{-\left(\frac{wd_{ji}}{\sigma}\right)^2} \quad (4)$$

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 :

$$wd_{ij} = \frac{d_{ij}}{w_{ij}} \quad (5)$$

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 .

Dynamic Role Analysis method Based on Weighted Directed Topological Potential

Any complex network is not static, since the characteristics of each node in the network will change and the role of the node will change correspondingly. A node's role is not only determined by the performance of a short period of time, but depends on the overall development trend of the

network, which is the stable performance of nodes in a fairly long period of time. On the other hand, if the network is considered as a static network for a long time, it is obvious that the role will lose the change of time, and the role of the role is not accurate.

In this paper, we propose to analyze the role of nodes in a dynamic network, considering the variation of the topology of each node in a time series. We divide a longer period of time(T) into a series of time fragments(t), calculate the out-degree and in-degree topology potential of all nodes in each time fragment t, analyze the roles of nodes in t using the static analysis method and last consider the change of roles in all time fragments. We can find that the nodes' roles are varies with time.

First, we discuss the role analysis in time fragment t. 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 wd_{ij} .

(2) Discuss the value taking for parameter σ : determine the σ value that minimizes H according to Eq.2. 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 : $\varphi_{out}(v_i)$ and $\varphi_{in}(v_i)$.

(4) Role classification in time fragment t: define the typical roles in the area with $\varphi_{out}(v_i) \in [\varphi_{out}(v_i)_{\min}, \varphi_{out}(v_i)_{\max}]$ and $\varphi_{in}(v_i) \in [\varphi_{in}(v_i)_{\min}, \varphi_{in}(v_i)_{\max}]$:

$A(\varphi_{out}(v_i)_{\min}, \varphi_{in}(v_i)_{\min}), B(\varphi_{out}(v_i)_{\max}, \varphi_{in}(v_i)_{\min}),$

$C(\varphi_{out}(v_i)_{\max}, \varphi_{in}(v_i)_{\max}), D(\varphi_{out}(v_i)_{\min}, \varphi_{in}(v_i)_{\max}).$

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.

(5) Role analysis in time T: define the role vector in time T as $R=(r_{t1}, r_{t2}, \dots, r_{tn})$. r_{ti} ($i=1, 2, \dots, n$) is the role of the node in time fragment t_i . Dynamic role analysis is to analyze the law of all nodes' role vector, so that we might understand the characteristics of the nodes more comprehensively. Role analysis can use the general clustering method to identify different types of roles which have similar change rules in time T.

Experiment

Data set.

The data set includes all the communication data of 334 IP nodes within 60 hours, and the data structure is obtained as table 1 after processing.

Tab.1. data structure

SRCIP	DSTIP	SRC PORT	DST PORT	FILE LEN
10.24.248.253	10.54.240.5	445	3822	1765
.....

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.

We define m_i as:

$$m_i = \frac{p_i}{P} \times \frac{\text{len}_i}{LEN} \quad (6)$$

p_i represents the total number of intrinsic ports of node v_i , P represents the total number of all IP ports, len_i is the sum of in-and-out traffic of node v_i , and 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.

Result.

According to the above formulas, the in-degree and out-degree topological potential are calculated with the optimal σ value 1.405. T=60h, so we divide T into 60 fragments $t_1=t_2=\dots=t_n=1$ h. Calculate the roles of all nodes in each time fragment t, and the result taken one hour as an example is shown in Figure 1.

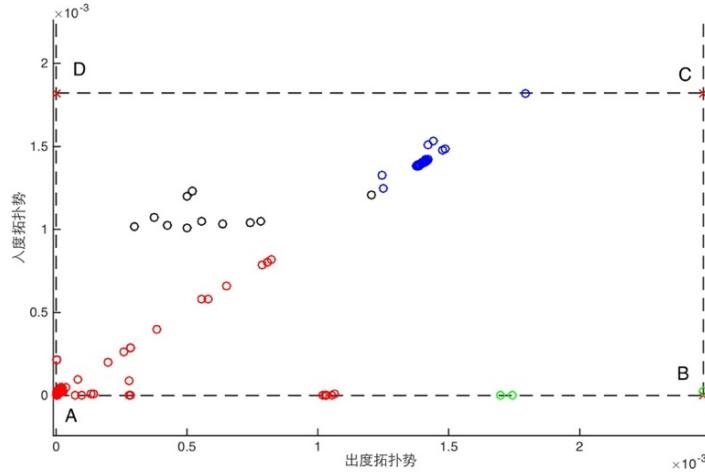


Fig.1.in and out degree topological potential of nodes and role partition result

Figure 1 shows the topological potential, four typical roles (A, B, C, D) and the final recognition of roles. In this 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.

Then four roles are represented by four colors:

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

(2) The main sending node: the green nodes in 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.

(3) The main transformation node: the blue nodes in 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.

(4) The main receiving node: the black nodes in 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.

Calculate R_1, R_2, \dots, R_n , and show the role-changes of every node in Figure 3. From top to bottom, each line expresses the roles varies with time of a node. From right to left, each column expresses a time fragment t. Role 1(blue), role 2(red), role 3(orange), role 4(green) correspond to the above analysis of the four types of roles.

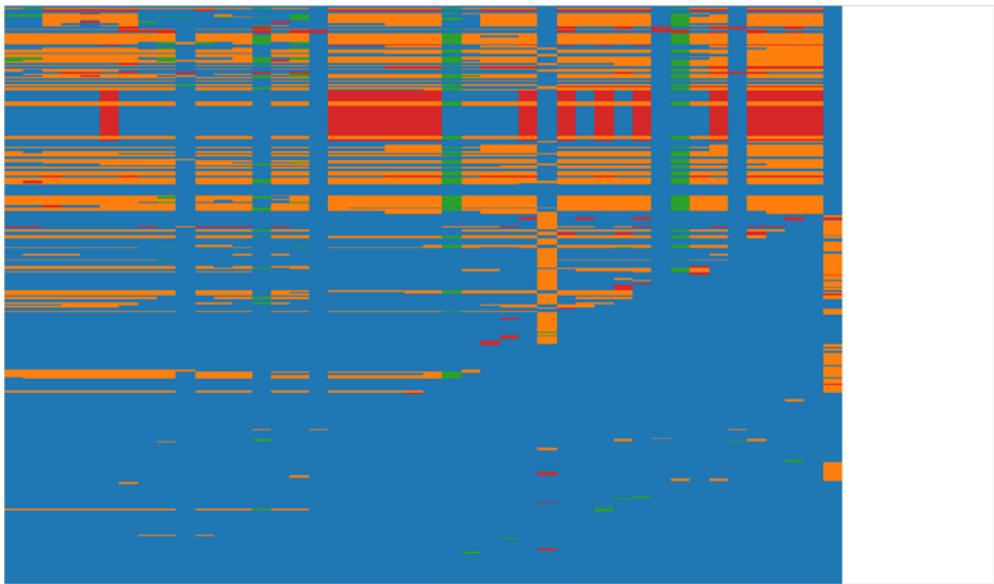


Fig.2. roles of 334 nodes in each time fragment

We find in Figure 2 that the roles of some nodes are simplex and stable, which means these nodes present only one certain role; the roles of some nodes appear alternately, which shows the circularity; and the roles of some other nodes change suddenly, which means these nodes belong to one role for almost the whole time but turn into another role at some point. According to these role-changes of nodes in time series, we cluster all the nodes and show the result as follows:

(1) Main service nodes: this class of nodes has shown a service role in almost the whole period of time, that is, the main transformation node, or the main receiving node, or the main sending node. These nodes bear almost all of the communication of the network, which is a very important class of nodes in the network.

(2) Unactive nodes: the class of nodes which with small in and out degree topological potential in the whole period of time is not active.

(3) Role alternation node: this kind of nodes is shown as a number of characters, and the characters appear alternatively or regularly.

(4) Mutation node: the class of nodes has been in a state of one role, but at some point mutation turns into the other part.

Conclusion

This research aimed at the limitation and shortage of role analysis in complex network, and put forward the dynamic role analysis method based on node directed weighted topological potential. In this part, We put forward the role analysis index of nodes with the network structure attribute and the node attribute, and divide the roles from the point of the node potential difference. At last, this paper highlights the dynamic role analysis method based on role vectors. This proposed method take into consideration the regularity and pattern of role-change, so that the result is more appropriate than usual analysis.

References

- [1] B.X. Li, M.J. Li, Networks model of the East Turkistan terrorism[J], Physica. A. 2015(419):479-486.
- [2] Z Li, DY Sun, B Li, ZH Lin, Detecting Changes in Terrorism Networks: A Perspective of Multi-Meta-Network Modeling [C], EISIC.2015.
- [3] C Lin, Social Network Mining Research Under the WEB Environment [D], Fudan University 2009.
- [4] E Estrada, JA Rodriguez-Velazquez, Subgraph Centrality in complex network[J], Phys Rev E,

2005(71):122-133.

- [5] LL Ma, C Ma, HF Zhang, BH Wang, Identifying Influential Spreaders in Complex Networks Based on Gravity Formula [J], *Physica A*.2015(451):205-212.
- [6] F Morone, HA Makse, Influence Maximization in Complex Networks Through Optimal Percolation [J], *Nature*, 2015.
- [7] LY Lv, Link Prediction on Complex Networks[J], *Journal of University of Electronic Science and Technology of China*. 2010(9):5-39.
- [8] T Zhu, Node Role and Group Evolution Research in Social Network [D], Beijing University of Posts and Telecommunications. 2011.
- [9] LIU Zhi-ming, LIIU Lu, Recognition and Analysis of Opinion Leaders in Microblog Public Opinions[J], *Systems Engineering*.2011,29(6):8-16
- [10] DUAN Song-qing, YU Xing-long, WU Bin, Method for Tracking the Role-based Evolution in Social Network[J], *Journal of Hunan University*.2015,42(8):132-140
- [11] Rossi, Modeling Dynamic Behavior in Large Evolving Graphs[C]//Proceedings of the 6th ACM International Conference on Web Search and Data Mining, Rome, Italy, Feb 4-8, 2013. New York, NY, USA:ACM, 2013: 667-676.
- [12] LI Yan-mei, LI Chuan, TANG Changjie, and so on.Role Evolving Outliers Detection in Dynamic Information Networks[J],*Journal of Frontiers of Computer Science and Technology*,2015(3):321-329
- [13] Gan Wen-Yan, Li DY, Wang JM. A hierarchical clustering method based on data fields. *Chinese Journal of Electronics*, 2006,34(2):258–262
- [14] Gan Wen-Yan, HE Nan, Li De-Yi.Community Discovery Method in Networks Based on Topological Potential[J],*Journal of Software*, 2009,20(8):2241-2254