2012 International Conference on Innovation and Information Management (ICIIM 2012)
IPCSIT vol. 36 (2012) © (2012) IACSIT Press, Singapore

Evaluation of Node Importance in Public Opinion Based on Topological Potential

SUN Rui[1,a], LUO Wan-Bo[1,b] and LI Lin[1,c]

[1]School of Computer Science, Sichuan University, Chengdu, China

[a]cug123456@126.com, [b]wbluo@scu.edu.cn, [c]lilin@scu.edu.cn

Abstract: The evaluation of node importance is the main research direction in public opinion field, which is much significant to accurately find out the influent nodes for the propagation and evolution of public opinion, furthermore to effective control and predict public opinion situation and in-time guide it. On the basis of numerous research of scholars and combined by the topology structure of network and the attributes of node itself, this paper introduces the idea of data field in theoretical physics and establishes the evaluation model of node importance in public opinion based on topological potential. Through the theoretical and experimental analysis, it is proved that this method can evaluate the importance of nodes in a fast and accurate way in propagation network of public opinion, which is significant both to theory and practice.

Keywords: Complex Networks; Node Importance; Public Opinion; Topological Potential.

1. Introduction

In short, the public opinion is common opinion of public on social phenomenon and social issues. With the rapid development of information technology, Internet has been recognized as the "fourth media" after newspapers, radio and TV, it become a main carrier of public opinion spread [1]. Due to the Internet has fast, flexible merit and other technology characteristics in network transmission, people pay more attention to the research on the spread, monitoring, warning and situation analysis about public opinion in networks.

The statistical research found that public opinion in networks has small-world effect [2], scale-free property [3] and community structure [4], so the related models of complex network become the main research methods on public opinion. The subject of public opinion is people, also is the participants of topics, the research on behavior and function of every individual in the process of public opinion formation and evolution has important significance. Relative to the complex network, that is quickly and accurately to find the important nodes.

The evaluation methods of node importance are essentially derived from graph theory and graph-based data mining, focus on the measure about network topology. Social network analysis approaches include Degree, Closeness, Betweeness, Information, Eigenvector and Cumulated nomination, etc. The system science research approach use the connectivity of network to reflect the integrity of certain functions, by measuring the extent of destroyed connectivity of network which caused by removing the nodes to reflect the importance of nodes [5]. The basis of the research is the core and coritivity theory [6]. In the Internet search field, some methods such as PageRank [7] and HITS [8] not only think of themselves, but also consider the importance of neighbors.

In the real-world networks, nodes often are not a point on mathematical sense, but are the entities which have many attributes. Especially in propagation network of public opinion, as a node the individual is obviously different, it has the attributes such as the extent of interested in a subject, influenced by others, the activity etc. To some extent, these attributes become the main driver of evolution even more than the network topology. Therefore, the investigation of node's attributes is the important methods on node importance [9, 10].

This paper introduces the idea of data field based on complex networks and establishes the evaluation model of node importance based on topological potential in public opinion. Through the theoretical and experimental analysis, it is proved that this method can evaluate node importance in an accurate way, which is significant both to theory and practice.

2. Evaluation of Node Importance Based on Topological Potential

Topological Potential of Nodes

The topological potential of nodes based on the concept of data field from cognitive physics [11]. British physicist Michael Faraday first proposed the concept of potential field, which was used to describe non-contact interactions between matter particles. Taking into account the character of complex networks, we adopt the Gaussian potential function to describe the interaction between nodes, which represent short-range field [12] and has a good mathematical property, therefore the corresponding field is called the topological potential field [13]. The size of the topological potential of nodes describes the value of potential of any nodes, which is effected by itself and neighbors in the network. In the network topology, we use the topological distance to compute the potential of nodes generated by passing through the network.

Definition 1: Given network $G=(V,E)$ with n nodes and m edges, where $V=\{v_1,v_2...,v_n\}$ denotes set of nodes, $E\subseteq V\times V$ denotes set of edges, and $|E|=m$. According to the potential function definition of data field [13], the topological potential of node $v_i\in V$ can be expressed as

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

Where d_{ij} denotes the distance between v_i and v_j , it is measured by the shortest distance. Impact factor σ control the effected area of each node; $M_j\geq 0$ denotes the quality of node $v_j\,(j=1,2,...,n)$, which is used to describe the attributes of node.

3. Evaluation Algorithm of Node Importance Based on Topological Potential

Given the propagation network of public opinion $G=(V,E)$, the description of evaluation algorithm of node importance based on topological potential as follows:

1. Computing node quality $M_j, j=1,2,...,n$

1) Building the attribute matrix

The basic attributes of public opinion participants can represents by recognition, debate, influence, obstinacy, activity and the extent which affected by mainstream media, officials, law and so on [9].

Therefore, we suppose that $X=\{x_{i1},x_{i2},...,x_{im}\}$ denotes m attributes of node v_i , where x_{ij} is the j-th attribute of the i-th node. When f_i denotes the attribute function, $x_{ij}=f_i(v_i)$, $i=1,2,...,n$, $j=1,2,...,m$. The attribute values of each node constitute the attribute matrix, where X_i , $i=1,2,...,m$ denotes m attributes. The attribute matrix is shown in table 1.

Table.1 Attribute matrix

	X_1	...	X_j	...	X_m
v_1	x_{11}	...	x_{1j}	...	x_{1m}
...	...	...	...	...	...
v_i	x_{i1}	...	x_{ij}	...	x_{im}
...	...	...	...	...	...
v_n	x_{n1}	...	x_{nj}	...	x_{nm}

Table.2 Attribute values

	Debate X_1	Influence X_2	Obstinacy X_3	Activity X_4
v_1	6	2	8	48

v_2	8	4	9	80
v_3	4	1	7	45
v_4	9	20	10	120
v_5	1	8	5	55
v_6	5	0	6	20
v_7	9	10	7	108
v_8	0	4	5	80
v_9	2	2	9	95
v_{10}	8	4	6	76

Therefore, the attribute value x_{ij} of v_i are defined as follows: x_{i1} denotes debate of v_i; x_{i2} denotes influence of v_i; x_{i3} denotes obstinacy of v_i; x_{i4} denotes activity of v_i.

2) Normalizing the node attribute value

Each element in the attribute matrix is measured by different methods, they have different dimension. In order to comparison and effective evaluation, we should normalize the attribute values and transform them to $[0,1]$.

We suppose the original attribute matrix is $X = \{x_{ij}\}$, the transformed attribute matrix is $Z = \{z_{ij}\}$, $i = 1,2,...,n$, $j = 1,2,...,m$, $x_j^{\max}$ and $x_j^{\min}$ are the maximum and minimum in the j-th column of matrix.

$$z_{ij} = f_i(v_i) = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \ . \tag{2}$$

According to Eq.2, we compute the normalized attribute value, the best value is 1, and the worst value is 0.

3) Computing node quality M_i

$$M_i = \frac{1}{m}\sum_{j=1}^{m} w_j z_{ij} \ , \ i = 1,2,...,n, \ \sum\nolimits_{i=1}^{m} w_i = 1 . \tag{3}$$

2. Computing the topological potential of each node $\varphi(v_i), i = 1,2,...,n$

Get M_i into Eq.1 to compute $\varphi(v_i)$.

3. Sorting node importance by comparing the value of topological potential

Sort node importance according to the size of topological potential, and the greater the topological potential, the more importance in network.

4. Parameter analysis

There are two parameters in the algorithm need to discuss: the weight of attributes in node quality and the impact factor σ in the formula of topological potential.

The weights of attributes reflect the disparity between attribute values and the contribution degree for node quality. The determination of the weights is always the key and difficult point in multiple attributes decision making problems. According to the actual situation, we usually adopt some data mining methods, such as the expert experience grade, data statistics analysis, genetic algorithm etc. This paper adopts the method presented by the literature [9], which compared to the attribute index in pairs and then solve equations to get a group of weight vector with the least square method.

We suppose that every node has m attributes, compare the importance of each attribute in pairs, so need to compare $C_m^2 = \frac{1}{2}m(m-1)$ times. b_{ij} denotes the relative importance from the i-th attribute to the j-th attribute, and we think that this is the ratio of the weight w_i of attribute i to the weight w_j of attribute j.

$$b_{ij} = \frac{w_i}{w_j} \text{ .} \tag{4}$$

The result of the comparison of m attributes in pairs constitute matrix B .

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix} = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \dots & w_1/w_m \\ w_2/w_1 & w_2/w_2 & \dots & w_2/w_m \\ \dots & \dots & \dots & \dots \\ w_m/w_1 & w_m/w_2 & \dots & w_m/w_m \end{bmatrix} . \tag{5}$$

Where $b_{ij} = \frac{1}{b_{ji}}, b_{ij} = b_{ik} \times b_{ki}, b_{ii} = 1$. (6)

and $\sum_{i=1}^{m} b_{ij} = \frac{\sum_{i=1}^{m} w_i}{w_j}$. (7)

When $\sum_{i=1}^{m} w_i = 1$, $w_j = \frac{1}{\sum_{i=1}^{m} b_{ij}}$. (8)

There are some subjective factors in estimate about the relative importance b_{ij} . We use least square method to solve the weights in order to avoid error.

Solving equations

$$\left.\begin{array}{l} \min \sum_{i=1}^{m} \sum_{j=1}^{m} (b_{ij} w_j - w_i)^2 \\ \sum_{i=1}^{m} w_i = 1 \\ w_i > 0, i = 1, 2, \dots, m \end{array}\right\} . \tag{9}$$

with Lagrange multiplier method to solve the binding scalar optimization problems. The Lagrange function [10] is $L = \sum_{i=1}^{m} \sum_{j=1}^{m} (b_{ij} w_j - w_i)^2 + 2\lambda \sum_{i=1}^{m} w_i - 1$, L solve partial derivative of $w_l (l = 1, 2, \dots, m)$, and make it zero, so

$$\sum_{i=1}^{m} (b_{il} - w_i) b_{il} - \sum_{j=1}^{m} (b_{lj} - w_l) + \lambda = 0, l = 1, 2, \dots, m . \tag{10}$$

Eq.10 and $\sum_{i=1}^{m} w_i = 1$ are $m+1$ equations, where $w_1, w_2, \dots, w_m$ and λ have $m+1$ variables, therefore the weights of attributes are $w = [w_1, w_2, \dots, w_m]^T$.

The impact factor σ in the formula of topological potential is used to control the effected area of each node. Usually we can optimize σ according to the potential entropy of nodes in network.

Definition 2: Given network $G = (V, E)$ and the topological potential field corresponding σ , make the potential values of $v_1, v_2, \dots, v_n$ are $\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n)$. The corresponding potential entropy of topological potential field is

$$H = -\sum_{i=1}^{n} \frac{\varphi(v_i)}{Z} \log\left(\frac{\varphi(v_i)}{Z}\right) \text{ .} \tag{11}$$

Where $Z = \sum_{i=1}^{n} \varphi(v_i)$ denotes normalized factor.

When $\sigma = 0$, $\varphi(j \to i) \to 0$, therefore $\varphi(i) = M_i = M$, potential entropy approach to $H_{\max} = \lg N$. When $\sigma \to \infty$, $\varphi(j \to i) \to M_j$, this means that there are same effect between any two nodes, $\varphi(i) = NM$, after the normalization by Z , potential entropy approach to $H_{\max} = \lg N$ again. The size of potential entropy reflect the strength of uncertainty [14], with the impact factor increasing monotonically, the curve of potential entropy reaches the maximum value at both ends, in the middle there will be a minimum. At this time the node potential has the most uneven distribution and the minimum uncertainty. The relationship between potential entropy H and impact factor σ is shown in figure 1.

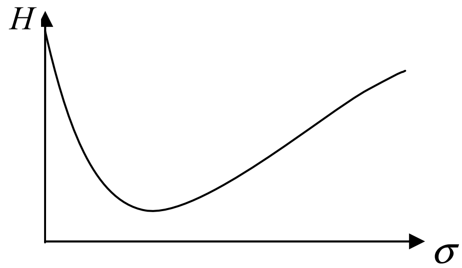

Fig. 1 Plot of potential entropy H vs. impact factor σ

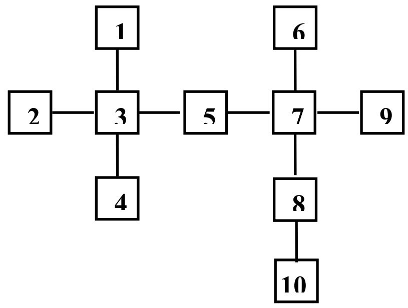

Fig.2 A simple opinion network

The optimization of parameter σ is a minimization problem of single variable nonlinear function $H(\sigma)$. Usually we can use simple test algorithm, random search algorithm and simulated annealing algorithm to solve. Considering the larger time cost of iterative calculation for large-scale complex network's topological potential, in the actual solution, we can first approximate estimate optimization interval of σ, and then search for the exact optimal value.

5. Experiment and analysis

As shown in figure 2, a propagation network of public opinion contains 10 nodes and 9 edges, where each node represents the participants in public opinion. Node's attribute values as shown in table 2.

Normalizing the node attribute value following Eq.1, the results are shown in table 3.

Table.3 The normalized attribute values

Relative importance degree	Definition	Explain
1	Equally important	Two attributes are equally important
3	Slightly important	A attribute slightly important than another
5	Quite important	A attribute important than another
7	Obvious important	A attribute obviously important than another
9	Absolutely important	A attribute deeply important than another
2,4,6,8	Intermediate values	Need to compromise

Table.4 The values set for judgment

	Debate Z_1	Influence Z_2	Obstinacy Z_3	Activity Z_4
v_1	0.6667	0.1000	0.6000	0.2800
v_2	0.8889	0.2000	0.8000	0.6000
v_3	0.4444	0.0500	0.4000	0.2500

	Debate Z_1	Influence Z_2	Obstinacy Z_3	Activity Z_4
v_4	1.0000	1.0000	1.0000	1.0000
v_5	0.1111	0.4000	0.0000	0.3500
v_6	0.5556	0.0000	0.2000	0.0000
v_7	1.0000	0.5000	0.4000	0.8800
v_8	0.0000	0.2000	0.0000	0.6000
v_9	0.2222	0.1000	0.8000	0.7500
v_{10}	0.8889	0.2000	0.2000	0.5600

In order to compare the relative importance between attributes, we get b_{ij} with reference to the level table given by Saaty, which are shown in table 4.

According to the experience about public opinion research, we can think that the debate X_1 compared with the influence X_2 between "Equally important" and "Slightly important", therefore $b_{12} = 2$. Without loss of generality, this paper compare the importance of each attribute in pairs and according to Eq.4 get the attribute relative importance judgment matrix B [9]:

$$B = \begin{bmatrix} 1 & 2 & 3 & 7 \\ 1/2 & 1 & 1 & 5 \\ 1/3 & 1 & 1 & 5 \\ 1/7 & 1/5 & 1/5 & 1 \end{bmatrix}$$

According to Eq.10 and $\sum_{i=1}^{4} w_i = 1$, the weights of attributes are $w = [0.5037, 0.2461, 0.1922, 0.0579]^T$.

Then we compute each node quality according to Eq.3. The results are as follows: $M_1 = 0.1230$, $M_2 = 0.1714$, $M_3 = 0.0980$, $M_4 = 0.2500$, $M_5 = 0.0437$, $M_6 = 0.0796$, $M_7 = 0.1886$, $M_8 = 0.0210$, $M_9 = 0.0834$, $M_{10} = 0.1420$.

According to the discussion about parameter optimization on section 2, we choose $\sigma = 1.374$.

Therefore according to the Eq.1, we compute topological potential $\varphi(v_i)$. The comparison between some evaluation methods and topological potential are shown in table 5.

Table.5 The evaluation values of node importance

Node	Degree	Betweenness	Closeness	Simplified Topological Potential	Topological Potential
v_1	1	0	0.037	0.192	0.238
v_2	1	0	0.037	0.192	0.281
v_3	4	0.468	0.053	0.344	0.469
v_4	1	0	0.037	0.192	0.350
v_5	2	0.444	0.059	0.283	0.301
v_6	1	0	0.040	0.193	0.211
v_7	4	0.556	0.059	0.356	0.356
v_8	2	0.178	0.044	0.250	0.241
v_9	1	0	0.040	0.193	0.214
v_{10}	1	0	0.032	0.171	0.179

From the experiment results, we can reach a conclusion that this method presented in this paper can more accurate shows the node's position information and difference in network. Comprehensive consideration for the node's properties and the network topology will more close to the actual situation than pure consideration on the network topology. We find that from the experiment results, node v_7 is the most important node by using some traditional evaluation methods, but if we still consider the comprehensive attributes of nodes, node v_4 more important than node v_7. Node v_1, v_2 and v_4, node v_6 and v_9 seem to be completely symmetrical structure in the network topology, so they will also have the same importance according to the evaluation indexes of topological structure. But in the public opinion network the nodes represent the participants of topic, they have differences take for granted, which abstract perform to be the different attributes of nodes. Therefore, node v_4 is more important than node v_2, node v_2 is more important than node v_1, and node v_9 is more important than node v_6.

Total time complexity of algorithm depends on the calculation of topological potential, it is $O(n^2)$ in the worst case and $O(m+n^{3/\gamma})$ in the best case, where n nodes, m edges and $2<\gamma<3$ denotes a constant.

6. Summary

The research about node importance in public opinion is helpful to reasonable and effective monitor and forecast the development and change of public opinion, exact find out those nodes which have important influence to the spread of public opinion in the complex environment. The monitoring and technical processing of important nodes can very effective guide public opinion, at the same time they have the important meaning to create a healthy and orderly network environment for public opinion and the harmonious society. This paper summarizes the research about evaluation of node importance, references the idea of data field in physics and establishes the evaluation model of node importance in public opinion network. This method takes into account the network topology and node's attributes, so it can more accurate evaluate node importance. Through the theoretical and experimental analysis, it is proved that this method is significant both to theory and practice.

7. References

[1] CHEN Wei-xing: Network communication and social development (China Broadcasting and Television Press, Beijing 2001).

[2] D.J. Watts and S.H. Strogatz: Collective dynamics of small world networks, Nature, Vol. 393 (1998), p. 440-442.

[3] A.L. Barabási and R. Albert: Emergence of scaling in random networks, Science, Vol. 286 (1999), p. 509-512.

[4] M. Girvan and M.E.J Newman: *Community structure in social and biological networks*, Proc. of the National Academy of Sciences of the United States of America, Vol. 99 (2002), p. 7821-7826.

[5] TAN Yue-jin, WU Jun, DENG Hong-zhong: *Evaluation Method for Node Importance based on Node Contraction in Complex Networks*, Systems Engineering -Theory & Practice,26,11,(2006).

[6] XU Jin, XI You-min and WANG Ying-luo: *On system core and coritivity*, J. Sys. Sci. & Math. Scis., 13(2), 102-110,(1993).

[7] S. Brin and L. Page: The Anatomy of a Large-scale Hypertextual Web Search Engine, Computer Networks, 30,107-117,(1998).

[8] J.M. Kleinberg: *Authoritative Sources in a Hyperlinked Environment*, Journal of the ACM, 46, 5,(1999).

[9] HU Chao-lang, WU Rong-jun and ZHOU An-min: Study on the Formation Models of Network Public Opinion Based on Incline Degrees of Opinions of Agents, Journal of Sichuan University (Engineering Science Edition), Vol. 41,(2009), p. 196-201.

[10] WU Qing-feng, KONG Ling- jiang and LIU Mu-ren: Influence of person's character upon the evolution of the cellular automata model for public opinion, Journal of Guang Xi Normal University, Vol. 22 (2004), p. 5-9.

[11] LI De-yi: *Artificial Intelligence with Uncertainty*, NewYork: CRC Press. 207-267,(2007).

[12] W.Y. Gan: *Study on the clustering problem in data mining* (University of Science and Technology, Nanjing 2003).

[13] GAN Wen-yan, HE Nan, LI De-yi and WANG Jian-min: Community Discovery Method in Networks Based on Topological Potential, Institute of Software, 20, 2241-2254,(2009).

[14] XIAO Li-ping, MENG Hui and LI De-yi: Approach to Node Ranking in a Network Based on Topology Potential, Geomatics and Information Science of Wuhan University, 33, 379-383,(2008).