A New Social Network Partition Method

Huang Jian-Heng1,2, Yu Jun-Qi2, Zhao Min-Hua2

1 School of Civil Engineering, Xi’an University of Architecture & Technology, Xi’an, China
2 School of Information & Control Engineering, Xi’an University of Architecture & Technology, Xi’an, China
E-mail: Jhhxian@126.com

Abstract. In allusion to issues that partitioning algorithm of traditional social network community is generally lacking in comprehensive consideration of node and link attributes as well as full expression of model and mechanism using attributes of node and link. This paper presents a social network community partitioning algorithm fusing node and link attributes. The algorithm has blended similarity of node properties, link weights between nodes and link attribute information, making definition of the similar weights, and on this basis, social networking community division is realized combined with condensation algorithms. Experiments show that the algorithm has remarkable effects on community partition with obvious attributes in social network.

Keywords: Social network; Social network partition; Centrality; Controllable routing

1 Introduction

With the rapid socio-economic development, people have conducted increasingly frequent exchanges and a variety of social activities related to work and life, and information exchange is also developing rapidly. In the process of information exchange, each person's social attributes, including social relations, social behavior, etc., can be shown explicitly or implicitly, gradually forming social networking such as instant messaging chat networks, telecommunication networks and microblog networks. The study has found that social networking prevails in "community structure". How to divide the community structure is of great significance to the study on social networks.
2 Model and Related Definitions

2.1 Social Networking Model and Definitions

Social networks are generally indicated by informative graph, and simple social network is shown in Figure 1:

Fig. 1. Simple social network model

Definition 1. Informative graph. Assuming \( G=(V,E) \) in the diagram, in which \( V \) is the individual (or node) in network. There is a total of \( n \) nodes in \( V=\{v_1,v_2,\ldots,v_n\} \). \( E \) is the set of node link (or edge) in the network, \( E=\{e_1,e_2,\ldots,e_m\} \), a total of \( m \) edges. There are \( P \) attributes in individual, and attribute set of node \( V \) is \( V_{\text{attr}}=\{a_1,a_2,\ldots,a_p\} \). There is \( h \) attributes in link in total, and the attribute set of link \( E \) is \( E_{\text{attr}}=\{b_1,b_2,\ldots,b_h\} \). From the above information, properties matrix of node and link in the graph can be built. Link matrix of informative graph can be expressed by the following matrix:

\[
\begin{pmatrix}
A_{i1} & \cdots & \cdots & A_{in} \\
\vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
A_{ni} & \cdots & \cdots & A_{nn}
\end{pmatrix}
\]

Among them, \( A_{ij} \) is the link weight from node \( i \) to node \( j \), and \( A_{ji} \) is the link weight from node \( j \) to node \( i \). For powerless undirected graph, element value in the matrix is 0 or 1, and \( A_{ij}=A_{ji} \).
2.2 Information gain

Information gain is defined as the difference between the original information needs and new needs [13]. The original information needs refers to the desired information of original sample classification. The new requirements mean information needs of sample classification based on the partition of known property A.

Supposed that there are m sample categories in D, and the ratio of the ith category \( (i = 1, 2, ..., m) \) to total number of samples is \( p_i \), then:

\[
Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)
\]  

(1)

Supposed that the samples in D are divided according to characteristics of A, D can be divided into v subsets with A, in which the sample in \( D_j \) has value \( a_j \) in terms of the A. According to this division of A division, the desired information required in sample partition of D is:

\[
Info_A(D) = \sum_{j=1}^{v} \left( \frac{|D_j|}{|D|} \cdot Info(D_j) \right)
\]  

(2)

The information gain of attribute A is:

\[
Gain(A) = Info(D) - Info_A(D)
\]  

(3)

Among it, \( Gain(A) \) is information gain of attribute A; \( Info(D) \) is original information need, and \( Info_A(D) \) is new information need.

2.3 Modularity

Assumed that \( G = (V, E) \) in the graph, there is a total of \( n \) nodes and \( m \) edges, and in powerless undirected graph, modularity Q on the node is defined [6] as:

\[
Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j)
\]  

(4)
Among it, $k_i = \sum_{j=1}^{n} A_{ij}$ is the number of links (that is, the degree of $V_i$) owned by $V_i$. $C_i$ is the community that node $i$ belongs to. When the node $i$ and node $j$ is in the same community, $\delta(C_i, C_j) = 1$, otherwise it is zero. The role of module is mainly to evaluate the stand or fall of community partition and the strength of the network community structure. $Q$ value is in the range of $(0, 1)$. The greater value means better result of the division and more obvious community structure. For networks with modularity usually from 0.3 to 0.7, their communities’ structures are more evident.

2.4 Normalized Mutual Information

NMI is used to measure the difference between the divided community and real community. Assumed that the real community is divided into $C_o$, and the community obtained through algorithm is divided into $C_e$. NMI is defined as:

$$NMI(C_o, C_e) = \frac{H(C_o) + H(C_e) - H(C_o, C_e)}{\sqrt{H(C_o)H(C_e)}}$$

Among it, $H(C)$ represents the Shannon entropy of divided $C$. When $C$ and $C_o$ are fully consistent, $NMI(C_o, C_e) = 1$; when $C_e$ and $C_o$ are completely different, $NMI(C_o, C_e) = 0$. When $NMI(C_o, C_e)$ is between $(0, 1)$, the greater value indicates that the divided result is closer to that of the real community.

3 Summary and Outlook

As for the community discovery of social networking, this paper has proposed community partitioning algorithm based on similar weights integrated with attributes, making appropriate improvements of modularity combined with powerful and directed network, and has applied this algorithm in community division issues of real social networking. With similar weight as the basis for dividing the community, the divided communities have higher modularity. The algorithm has commendable applicability and effectiveness, with more obvious effects particularly for network that node attribute information has greater impact on community partition. There has been no in-depth study on what to think when there are excessive node attributes and on the coefficient setting of link attributes in the algorithm currently, and in addition, the considerations of weight setting for link attributes are too simple. Issues on how to divide into more real and effective results and to improve the NMI value all need in-depth study in the next step.
Acknowledgments. This work was supported by the project of the Ministry of Housing and Urban-Rural Development (2012-K1-1). The scientific research project of department of education of Shaanxi province (12JK0999).

Reference