2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications

Evaluating User Community Influence in Online Social Networks

Zhengli Xiong, Wenjun Jiang, and Guojun Wang*

School of Information Science and Engineering
Central South University
Changsha, Hunan Province, P. R. China, 410083
*Correspondence to: csgjwang@mail.csu.edu.cn

***Abstract*—Word-of-mouth (or viral) marketing effects have raised the problem of how to discover the influential members in a community in online social networks (OSNs). In this paper, we develop a novel model to evaluate the influential strength and identify the most influential users of a community in OSNs. We define a new concept of community label, which is used to represent the attributes that are possessed by most of the members in a community. By analyzing the interpersonal structure and the unique characteristics of a community in OSNs, we show that the number of friends, the quality of friends, and the community label, are three key factors to a user's community influence which means a user's social influence in a community. Experimental results show that our model can evaluate a user's community influence with higher efficiency and more rationality than traditional models.**

***Keywords*-User Community Influence; Online Social Networks; Community Label**

I. Introduction

The Internet has dramatically changed the way that people communicate with each other. In the past few years, online social networks (OSNs) have gained significant popularity. They have become pervasive platforms that focus on building social networks or social relations among people. Sathik [1] defined a cyber community in an OSN as a group of people who share common interests such as hobbies or occupations, and interact with each other through the OSN. Communities can provide valuable information or resources. For example, they can represent the sociology of the web, and their study about user influence can provide convenience for marketing.

With the advent of communities in OSNs, a user's social influence in a community, coined as user community influence (UCI for short) in this paper, is increasingly used as the key point in the word-of-month (or viral) marketing for sales and advertising. The essence of word-of-mouth marketing is to reach out to a broad set of influential customers and attract considerable attention via community interactions [2]. Recommending new products to those influential customers can lead to a series of influence: some of them will recommend the products to their friends. Through word-of-mouth diffusion, more people will be influenced and buy the new products.

As an emerging social media that exploits influential users of communities in OSNs, a community system has gained a lot of attention. It has become a very good platform to promote new products and services for a company. However, in order to get a larger influential range with less cost, it is important to identify the most influential users.

A common approach to identifying influential users is to analyze the social network structures of OSNs [3]. Opsahl et al. [4] presented an evaluation model of a user's reputation based on degree centrality. Newman [5] discovered the very notion of influential users that is closely related to closeness centrality. Katona et al. [6] presented an evaluation model of a user's reputation using betweenness centrality. In this paper, we extract three factors that are important to a user's social influence in a community: the number of his friends, the quality of his friends, and the community label. All these factors are considered to develop a more comprehensive and robust model to evaluate a user's social influence in a community.

Based on the PageRank algorithm [10], which is used to evaluate the importance of a web page, we propose a novel model to evaluate a user's community influence using community analysis. Our contributions are as follows:

1) We explore the factors of a user's social influence in a community in OSNs. By analyzing the interpersonal structure and the unique characteristics of a community in OSNs, we show that the number of friends, the quality of friends, and the community label, are the three key factors of a user's social influence.
2) Based on Markov chain [16], we propose a user's community influence propagation model to evaluate a user's social influence in a community.
3) We evaluate our model using the *arXiv* dataset [19]. Experimental results show that our model can evaluate a user's UCI with higher efficiency and more rationality than existing work that is based on the structures of social networks.

Organization. First, we introduce some related work in Section II. Next, we give preliminaries and problem statement in Section III. Then, we outline our user's UCI analysis in Section IV, and provide a markov representation of UCI in Section V. Experimental results are shown in Section VI. Finally, we conclude this paper in Section VII.

978-0-7695-4745-9/12 $26.00 © 2012 IEEE
DOI 10.1109/TrustCom.2012.152

II. Related work

There has been a lot of research on social influence analysis. Existing research suggested several influencing factors based on the structures of social networks, in which the most prominent three are degree, closeness and betweenness [3].

A. User Degree Centrality

User degree centrality (UDC) is a basic indicator and often used as the first step when studying networks [4]. It relies on the number of nodes that a focal node is connected to, and measures the involvement of the focal node in the network. Its simplicity is an advantage: only the local structure around a node must be known for it to be calculated. To formally describe this measure and ease the comparison among the different measures introduced in this paper, we formalize this measure as:

$$UDC_u = \frac{\sum_{v} X_{uv}}{N-1} \tag{1}$$

where u is the focal node, v represents all other nodes, N is the total number of nodes, and X is the adjacency matrix, in which the cell X_{uv} is 1 if node u is connected to node v, and 0 otherwise.

B. User Closeness Centrality

User closeness centrality (UCC) relies on the length of the shortest paths from a focal node to all the other nodes in the network, and is defined as the inverse total length [5]. The smaller the value of UCC, the more centrality of the user. It can be regarded as a measure of how long it will take to spread information from the focal node to all the other nodes sequentially. Newman [5] asserted that closeness was:

$$UCC_u = \left[\sum_{v} d(u,v)\right]^{-1} \tag{2}$$

where v represents all the other nodes and $d(u,v)$ is the shortest distance between u and v.

C. User Betweenness Centrality

User betweenness centrality (UBC) relies on the identification of the shortest paths, and measures the number of them that pass through a node [6]. A larger number indicates more centrality. It was introduced as a measure for quantifying the control of a user on the communication between other users in a social network. Betweenness centrality is defined as:

$$UBC_u = \frac{sp_{ij}(u)}{sp_{ij}} \tag{3}$$

where sp_{ij} is the number of shortest paths between two nodes, and $sp_{ij}(u)$ is the number of those paths that go through node u.

In traditional areas of social network analysis, the centrality of a node is an important indicator to evaluate the user's social influence. Existing research is usually based on structural properties, and mainly focuses on the static networks [7]. However, many networks, such as social networks, are not static objects; however, they evolve continuously. Dynamic network analysis has become popular [8], [9], in which user behavior and friendship among users were taken into consideration.

Tantipathananandh et al. [8] developed a framework for detecting dynamic community structure which focused on how to model community formation and evolution based on user behavior.

Tang et al. [9] proposed an evolutionary clustering algorithm to identify communities in a dynamic multi-mode network by mining informative actor behavior features and actor interactions from social media.

Therefore, in OSNs, the information about user behavior and friendship among users is certainly valuable for dynamic network analysis, especially for identifying user roles, as it reflects the behavioral regularity of individuals. Based on the above analysis, we propose a novel model to evaluate user influence, in which we emphasize the effects of both the community and the friendship.

III. PRELIMINARIES

In this section, we analyze the similarity between the PageRank algorithm [10] and our work, and give the problem definition.

A. PageRank

By using the hyperlink structure of the web, PageRank gets the recommendation of a page from the inlinking page, which inlinks into that page. Since the inlinks from good pages should carry more weight than the inlinks from marginal pages, each webpage should be assigned an appropriate rank score, which measures the importance of the page. The PageRank algorithm was formulated by Google founders Larry Page and Sergey Brin as a basis for their search engine. After webpages are retrieved by robot crawlers, they are indexed and cataloged, and PageRank values are assigned prior to query time according to perceived importance. The importance of each page is determined by the links to that page. The importance of any page is increased with the number of sites which link to it. Thus, the rank $PR(s)$ of a given page s is given by:

$$PR(s) = C \sum_{i \in B_s}^{i} \frac{PR(i)}{N_i} \tag{4}$$

where C is a normalized constant, B_s represents all pages pointing to s, and N_i represents the number of outlinks from i. Therefore, the PageRank value $PR(s)$ of s is mainly dependent on the following three factors:

1) The number of pages pointing to s;
2) The PageRank value of every page in B_s;
3) The outdegree N_i of the page in B_s.

The PageRank equation computes the importance of pages in a Web graph relative to the quantity and quality of outlinks. To be globally relevant, the user's UCI also needs the analysis of the quantity and quality of friends. Inspired by the PageRank algorithm, we develop the initial community influence model.

B. Problem Definition

Formulism of Community in OSNs. A community is modeled as an undirected graph $G = (V, E)$, with V indicating the users in the community and E indicating the relationships between users. For example, in our experimental section, we study the co-authorship graphs where vertices represent authors of papers, and two vertices have an edge if the two corresponding authors have co-authored a paper.

We also denote the *Influence Ability* set by $W = \{W_{ij}|i,j \in V \; and \; <i,j>\in E\}$.

Problem Definition. In a given community $G = (V, E)$, due to the diversity of users, a user's UCI is various. Mathematically, a user's UCI can be formalized as a multivariate function $UCI_u = f(\theta_1, ...\theta_i, ...\theta_n)$, in which θ_i is the community influence factor, such as the number of friends, the quality of friends, and the community label. Our goal is to find an appropriate function to evaluate a user's social influence in a community in OSNs efficiently and effectively. Furthermore, a good evaluation model should not only distinguish the community influence of different users as much as possible according to our real social life, but also spread the information to achieve larger range in the community.

IV. USER COMMUNITY INFLUENCE ANALYSIS

In this section, we propose the user commnunity influence (UCI for short) model.

A. Solution Framework

Our solution framework is composed of five parts (as shown in Fig. 1).

1) We analyze the community influence factors of the number of friends, the quality of friends, and the community label, in which the quality of friends includes the *Attribute Similarity of Friends* and the community influence of friends.
2) We propose a heuristical model named *Influence Ability* based on the *Attribute Similarity of Friends*.
3) The initial community influence of a user is the summation of the user's influence abilities according to the number of friends.
4) We propose a community influence propagation model based on community label and influence of friends to reflect the dynamical information propagation.
5) Our evaluation model of a user's UCI is consistent with the construction model of the initial community influence and community influence propagation model.

B. Personal Community Influence Factors

Attribute Similarity of Friends. Online Social Networking is an important platform for sharing, organizing, and finding content and contacts, that is to say, they are a type of network which is surrounded by users' attributes (interests, hobbies et al.). Also, attribute similarity reflects the impact from one user to another [11]. Inspired by the idea of community detection based on node attribute similarity [12], we give the definition of the *Attribute Similarity of Friends*.

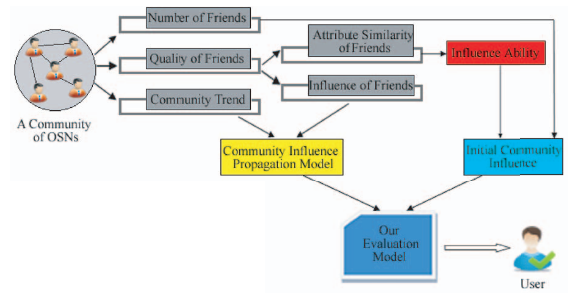

Fig. 1. Solution Framework.

Definition 1 (Attribute Similarity of Friends): *Attribute Similarity of Friends* is defined as the proportion of the number of the same attributes in all the attributes. For each nominal attribute a_k, if two connected nodes u, v have the same value, then increment the edge weight by 1. *Attribute Similarity of Friends* is defined as follows:

$$T_{uv} = \begin{cases} 0, & \text{u, v is not connected} \\ \frac{1}{A} \times \sum\limits_{k=1}^{A} \left(1 - |u.a_k - v.a_k|\right), & \text{u, v is connected} \end{cases} \tag{5}$$

where A is the total number of the user's attributes considered. If two nodes u, v are not connected, the T_{uv} is zero. For each user u, the larger T_{uv} is, the larger the ability that one can impact the other.

Furthermore, *"Things of a kind come together; birds of a feather flock together"*. In the real life, two persons who have closer relations will have more common friends. Therefore, in the scenario of a community, the more two users' common neighbors are, the larger the ability one can impact the other. Inspired by the idea of community detection based on common neighbor similarity [13], we propose the concept of *Influence Ability* to measure the ability one can impact the other based on the *Attribute Similarity of Friends* we defined above.

Definition 2 (Influence Ability): In a community $G = (V, E)$, the *Influence Ability* between u and v, denoted by W_{uv} is defined as:

$$W_{uv} = \frac{\sum\limits_{i=1}^{M} Min(T_{ui}, T_{vi})}{\sum\limits_{i=1}^{M} Max(T_{ui}, T_{vi})} \tag{6}$$

where M is the maximum of the number of the user u's or v's friends. Because the relationship is non-directional, W_{uv} is equal to W_{vu}. Intuitively, the larger the number of common close friends is, the larger *Influence Ability* is. This is coinciding with our real life.

Number of Friends. A user's UCI also depends on the number of friends. Generally speaking, the more friends the user has, the larger his UCI is. It is similar to PageRank, the more inlinks of a page has, the more important the page is.

Formally, the summation of the user's influence abilities with all his friends is the user's initial community influence.

Therefore, Based on *Influence Ability*, we propose the initial community influence in OSNs. It is defined as follows:

Definition 3 (Initial Community Influence): In a community $G = (V, E)$, the initial community influence of user u, denoted by UCI_u, is the total of his influence abilities, and his UCI is defined as follows:

$$UCI_u = \sum_{(u,v)\in E} W_{uv} \tag{7}$$

Furthermore, in the dynamic community in OSNs, user behavior can also impact community influence. In a real community where we live, there may exist communication between any two neighbors, that is to say, some information may flow from a certain user to his friends, which is a good way for advertisements. Based on the above analysis, we developed a novel model for the propagation of community influence in the next section.

C. User Community Influence Propagation Model

Community Label. *"Good companions have good influence while bad ones have bad influence"*. People are apt to be influenced by the majority. Also, according to the majority rule [14], we may predict the community label. We define the community label as:

$$c.a = \begin{cases} 1 \; more\, than\, half\, of\, users\, have\, attribute\, a \\ 0 \; else \end{cases} \tag{8}$$

Our basic intuition is, a user's opinions or recommendations are easier to be accepted by others in the community, if he has more attributes similar to community label. that is to say, he has the larger ability to impact others. Experimental results will validate our intuition. We call the similar degree between the user's attributes and the community labels as similarity with community.

Definition 4 (Similarity with Community (SC for short)): In a community $G = (V, E)$, a user's SC denoted by c_u is defined as:

$$C_u = \frac{1}{A} \times \sum_{k=1}^{A} (1 - |u.a_k - c.a_k|) \tag{9}$$

where A is the total number of all the attributes considered.

Influence of Friends. According to the *Matthew Effect* [15] in the social network, which is the phenomenon where "*the rich get richer and the poor get poorer*", the larger friends' UCI is, the larger his UCI is.

The community shown in Fig. 2(b) is evolved from the community shown in Fig. 2(a). In Fig. 2(a), c has only one close friend, but in Fig. 2(b), c adds two new friends, d and e. Under the condition of the same a, b and c, c's UCI become larger, then we claim that b's UCI becomes larger as a friend of c.

Based on above analysis, we propose the community influence propagation model.

Definition 5 (Community Influence Propagation Model): As we all know, a user's influence is from his friends. Therefore, in a community $G = (V, E)$, a user's UCI denoted by UCI_u is the total of influence contributions of the user's friends, and his UCI is defined as follows:

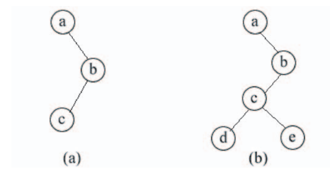

Fig. 2. User Network with Different Community Influence of Friends.

$$UCI_u = C_u \sum_{(u,v)\in E} UCI_{v\to u} \tag{10}$$

where $UCI_{v\to u}$ denotes the "contribution" to the node u's UCI. It reflects how much the attraction from node v to u.

However, how to measure the part community influence from node v to u? Let us present a simple star-like social network (Shown in Fig. 3).

Fig. 3 shows a network that consists of five nodes with node a being a center node. If node a advices node d to buy $iPhone4S$, d may buy because he has only friend a. If node d advices node a to do so, a may not buy it immediately before getting advice from his other three friends. That is to say, UCI_a is more important for d than UCI_d for a.

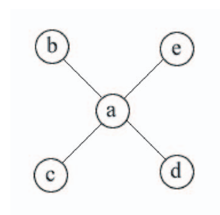

Fig. 3. A Simple Social Network.

According to above analysis, the contribution of a user's UCI should be provided by his friends. In Fig. 3, the contribution of a's UCI from his friend d is measured using $UCI_d * \frac{W_{ad}}{\sum_{(d,i)\in E} W_{di}}$, and the contribution of d's UCI from his friend a is measured using $UCI_a * \frac{W_{da}}{\sum_{(a,i)\in E} W_{ai}}$. Hence, Definition 5 is modified by Definition 6.

Our Model. *Definition 6 (User Community Influence Model)*: Based on PageRank, for a given user u, his UCI is re-defined as follows:

$$UCI_u = C_u \sum_{(u,v)\in E} UCI_v * \frac{W_{uv}}{\sum_{(v,i)\in E} W_{vi}} \tag{11}$$

where C_u is *Similarity with Community* of u, and W_{uv} is *Influence Ability* of u and v, v represents the friend of u, and i represents the friend of v. Therefore, the UCI value of u, that is UCI_u, is mainly dependent on the following three factors:

1) The number of friends;

2) The quality of friends, including influence of friends UCI_v and *Influence Ability* W_{uv};
3) The SC C_u.

V. MARKOV REPRESENTATION OF UCI

In the dynamic social network, a user's UCI is dependent on the contribution in the current state from his friends, and is only related to the number of friends, the quality of friends, and the community label, which are not changed during the propagation of the influence. This is very similar to Markov chain. Therefore, in this section, we use the property of markov chain to calculate the value of user's UCI.

A. Markov Chain

A Markov chain, named after *Andrey Markov* [16], is a mathematical system that undergoes transitions from one state to another, between a finite or countable number of possible states. It is a random process characterized as memoryless: the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessness" is called the Markov property [16].

We describe a Markov chain as follows: we have a countable set S, which is formed by possible states of S_i and is called *state space*, $S = \{S_1, ...S_i, ...S_n\}$, S_n is a conditional probability distribution over its historical state, which is determined by the initial distribution *(*P_1*,* P_2*, ...,* P_{n-1}*)*:

$$P(S_n = s|S_1, S_2, S_3, ..., S_{n-1}) = P(S_n = s|S_{n-1}) \quad (12)$$

Here, S is a certain state. Eq. (12) satisfies the Markov property.

The process starts from one state in S and transits successively from one state to another. Each transition is called a step. If the chain is currently in state S_i, then it moves to state S_j at the next step with a probability denoted by P_{ij}, and this probability does not depend upon which states the chain was before the current state. The probabilities P_{ij} are called *transition probabilities*. If *state space* is finite and constant, then the matrix of *transition probabilities* is defined as follows:

$$P_{ij} = P(S_{n+1} = j|S_n = i) = P \quad (13)$$

Here, for a discrete state space, given the chain's one-step transition matrix P, then $P^{(k)}$ refers to the *k-step* transition probability matrix which is the *k-power* of P.

B. Representation of UCI Using Markov Model

First, let us look at an example shown in Fig. 4. Fig. 4(a) shows that an initial social network is represented by an undirected graph in which users and social ties are represented by nodes and edges. If there is some information to be diffused, it will be evolved into the directed influential contribution network shown in Fig. 4(b). The arrows indicate the flow of the UCI. In the process of each flow, the UCI is always with the different contribution of flow to its connected node.

We adapt the transition matrix of Markov model to define our *Influence Ability* transition matrix P_a as follows:

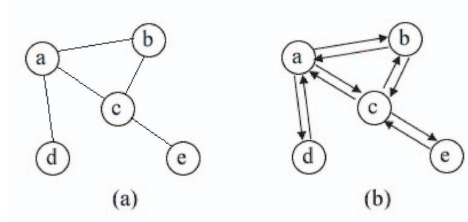

Fig. 4. (a) Initial Social Network (un-directed graph), (b) Influential Contribution Network (directed graph).

$$P_a = \begin{pmatrix} 0 & W_{21} & ... & W_{(n-1)1} & W_{n1} \\ W_{12} & 0 & ... & W_{(n-1)2} & W_{n2} \\ ... & ... & ... & ... & ... \\ W_{1(n-1)} & W_{2(n-1)} & ... & 0 & W_{n(n-1)} \\ W_{1n} & W_{2n} & ... & W_{(n-1)n} & 0 \end{pmatrix} \quad (14)$$

where the elements in P_a can be calculated according to Eq. (6). Intuitively, W_{uv} is equal to W_{vu}.

Furthermore, as we have mentioned before, a user's UCI can be influenced by his friends. In order to calculate a user's UCI, we re-define the apportion probability transition matrix P with P_a.

$$P = \begin{pmatrix} 0 & \frac{W_{21}}{W_2} & ... & \frac{W_{(n-1)1}}{W_{(n-1)}} & \frac{W_{n1}}{W_n} \\ \frac{W_{12}}{W_1} & 0 & ... & \frac{W_{(n-1)2}}{W_{(n-1)}} & \frac{W_{n2}}{W_n} \\ ... & ... & ... & ... & ... \\ \frac{W_{1(n-1)}}{W_1} & \frac{W_{2(n-1)}}{W_2} & ... & 0 & \frac{W_{n(n-1)}}{W_n} \\ \frac{W_{1n}}{W_1} & \frac{W_{2n}}{W_2} & ... & \frac{W_{(n-1)n}}{W_{(n-1)}} & 0 \end{pmatrix} * C \quad (15)$$

where W_u is the summation of the u's influence abilities. C is also a matrix with the elements using SC C_i. It is defined by:

$$C = \begin{pmatrix} C_1 & 0 & ... & 0 & 0 \\ 0 & C_2 & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & C_{n-1} & 0 \\ 0 & 0 & ... & 0 & C_n \end{pmatrix} \quad (16)$$

Suppose the initial distribution matrix of all users' UCI to be $UCI^{(0)} = (UCI_1, ...UCI_i, ...UCI_n)^T$, UCI_i represents the community influence of user u_i. and $UCI^{(n)}$ represents the distribution matrix of users' UCI after UCI being diffused by n steps. The $UCI^{(n)}$ is defined by:

$$UCI^{(n)} = P * UCI^{(n-1)} = ... = P^{(n)} * UCI^{(0)} \quad (17)$$

Also, because of the properties of Markov Model [17], matrix UCI satisfies the following equation:

$$\lim_{k \to \infty} P^{(k)} UCI^{(0)} = UCI \quad (18)$$

C. Convergence Analysis

In Eq. (18), $UCI^{(0)}$ is the initial distribution matrix of all users' UCI, if apportion transition probability matrix P is irreducible and non-periodic, then $\lim_{k\to\infty} P^{(k)}UCI^{(0)} = UCI$ is converged to stationary distribution UCI [17], and the UCI value of all users in our social network can be calculated [18].

We know that P_{ij} has a fixed value, since it is dependent on W_{ij} and SC, which is also related to the number and quality of the user's friends and community label. Hence, P_{ij} will not be changed during the iterations. When the UCI iterates, P will change monotonically as the iteration unfolding, and it will approach to either 0 or 1. Therefore, the UCI will converge to a reasonable value.

VI. EXPERIMENTAL EVALUATION

In order to evaluate our proposed model, we conduct a lot of experiments on a real-world social network. The results show that our model is efficient and accurate for the evaluation of a user's UCI.

A. Data Source

We use the dataset that is provided as a part of *the 2003 KDD Cup* [19], in which the collaboration network of scientists posted preprints on the condensed matter archive at *arXiv*. The social network is a paper co-operation network between the high-energy physicists called the HEP-TH (*high energy physics theory*), in which each node is used to represent an author and each edge is used to represent the cooperative relations between authors. This community network is an undirected graph. The statistics are shown in Table 1.

TABLE I
STATISTICAL INDICATORS OF ARXIV SOCIAL NETWORK

Items	Value
Number of nodes	31,163
Number of edges	120,029
Acquisition time	between January 1, 1995 and June 30, 2003

B. Experimental Design

We run our experiments on a 2.00 GHz dual core machine with 2G memory.

First, we calculate the number of friends for each node, which is as one of the major community influential factor.

Data Analysis. As mentioned above, the number of a user's friends is the major factor to affect the UCI. To figure out the relationship between the community influence and the number of friends, Fig. 5 shows the degree distribution of the community. As shown in the graph, the user degree is distributed exponentially. It fits with the power-law distribution [20], which has been found in many social networks. The relations are as follows:

- For the user sets A and B, if the number of friends of users in A is larger than that in B, then $|A|$ is smaller than $|B|$.

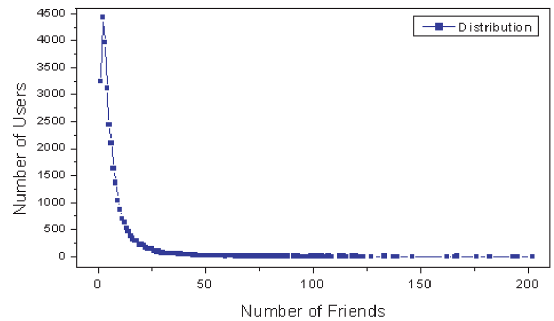

Fig. 5. Power-law Degree Distributions of Communities.

- In this community, the average number of a user's friends is around 8.

Furthermore, the experiment processes for our proposed model include calculating the UCI, UDC, UCC, UBC values, and the efficiency measure and the rationality measure. By analyzing the user's UCI value and the number of his friends, we know the relationship between the community influence and the number of friends. By analyzing the percentage of duplication calculated in different evaluation models, we can evaluate the efficiency of our model. By analyzing the range of the spread of influence calculated in different evaluation models, we can evaluate the rationality of our model.

C. Experimental Results

We conduct experiments based on different influence models and compare their corresponding results.

Analysis of User Community Influence. We calculate UCI, UDC, UCC and UBC values for each node in different influence models respectively.

The UCI value is shown in Fig. 6. We compare the experimental results and the data source, and conclude that the UCI value is related to the number of friends through the comparison with the original data source. The more the number of a user's friends is, the larger his UCI.

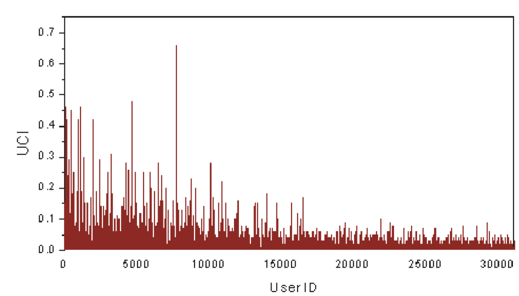

Fig. 6. The User's Community Influence.

Efficiency Measure. As we mentioned in problem definition, $UCI(u) = f(\theta_1, ...\theta_i, ...\theta_n)$ is a non-linear system, i.e., a user's UCI is diverse in our real-life world. Therefore, a good evaluation model should distinguish the user's UCI with f

among different users, i.e., avoiding duplication. To measure the effectiveness of our proposed model, the percentage of duplication η is defined as follows:

Definition 7 (Percentage of Duplication): Given a community $G = (V, E)$, the percentage of duplication is defined as follows:

$$\eta = \frac{\sum \partial(u,v)}{N}$$
$$\partial(u,v) = \begin{cases} 1 & M(u) = M(v) \\ 0 & M(u) \neq M(v) \end{cases} \quad u, v \in V \tag{19}$$

where M refers to the name of the evaluation model (such as UCI, UDC, UCC, UBC), and N is the total number of users in the community.

η represents the percentage of duplication. The evaluation model, in which η is smaller, can distinguish the user's UCI more easily, that is to say, it has the better evaluation effectiveness.

Property 1. The smaller the η is, the better the evaluation effectiveness is.

The comparison experiments on the *arXiv* dataset about η in different evaluation models are presented in Fig. 7. The results show that the percentage of duplication of our model is the lowest, compared with the three models. which indicates the high efficiency of our model.

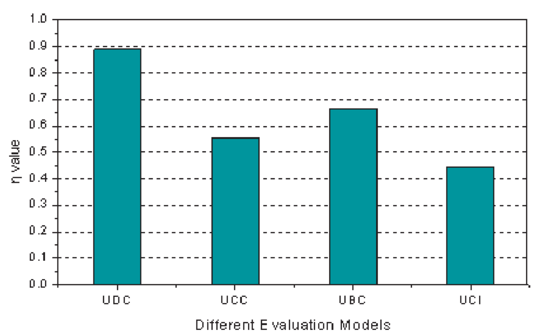

Fig. 7. Percentage of Duplication for Different Evaluation Models.

Rationality Measure. To measure the spread of influence of our proposed model, we use the fundamental propagation model, namely *Independent Cascade Model (IC)* [21]. The specific description of the IC model is as follows:

i) For the community G = (V, E), each node has only two states (active state and inactive state), and every node can only transit from inactive state to active state.
ii) At some time t, if a node u is active, then it will have the ability to activate its connected node v if v is in the inactive state. The probability that the node in the inactive state becomes active is denoted by p. If the influence value of the node u is larger than p, then it can activate its connected node v.
iii) If the node v is activated successfully, it will have the ability to activate its connected node.
iv) Repeat step ii and step iii for all nodes until there is no node that can be activated.

In the condition of the same total number of users, the number of users affected represents the range of the spread of influence. The evaluation model, in which the number of users are affected larger by the same initial seed, has the larger range of the spread of influence, that is to say, it has the better evaluation rationality.

Property 2. The larger the number of users affected by the same initial seed is, the better the evaluation rationality.

In our experiment by using *IC* in different probability p, the *Top-k* users with highest influence are selected as our seeds. We investigate the correlation of the number of users influenced by *Top-k* users and k under various models (shown in Fig. 8 and Fig. 9). As seen from these figures, our proposed model outperforms other models in terms of the maximization of the spread of influence. Our proposed model can significantly influence other users in its corresponding community.

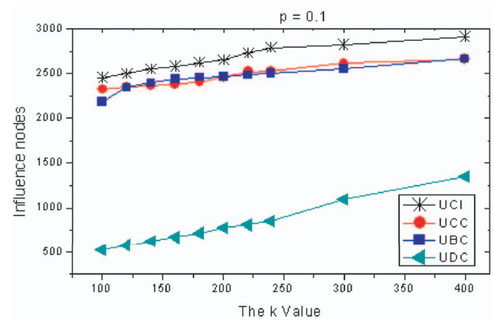

Fig. 8. Experimental results for Different Evaluation Models in the IC model with p = 0.1.

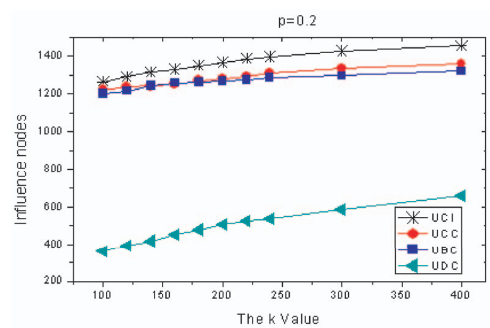

Fig. 9. Experimental results for Different Evaluation Models in the IC model with p = 0.2.

VII. CONCLUSION

We proposed a novel model to evaluate a user's social influence in a community in OSNs, in which we present two new concepts of *Influence Ability* and community label taking into account the three factors of the quantity of friends, the quality of friends, and the community label to construct the community influence. We also proposed the influence propagation model based on Markov chain. Experimental results show that our model has higher efficiency and more rationality than traditional structure based models.

In this paper, each attribute is given the same weight for simplicity. In fact, each attribute may have different influence and should be given different weight. Therefore, our future work is to decide the weight of different attributes and choose proper attributes to compute the attribute similarity between friends.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under grant numbers 61073037 and 61103035, and the Ministry of Education Fund for Doctoral Disciplines in Higher Education under grant number 20110162110043.

REFERENCES

[1] M. M. Sathik, K. S. Kannan, and A. Rasheed. Comparative Analysis of Community Discovery Methods in Social Networks. International Journal of Computer Applications, 14(8): 27-31, 2011.

[2] Y. M. Li, C. Y. Lai, and C. W. Chen. Discovering Influencers for Marketing in the Blogosphere. Information Sciences (Elsevier), 181(23): 5143-5157, 2011.

[3] T. Opsahl, F. Agneessens, and J. Skvoretz. Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. Social Networks (Elsevier), 32: 245-251, 2010.

[4] T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco. Prominence and Control: The Weighted Rich-Club Effect. Physical Review Letters, 101(16): 2008.

[5] M. E. J. Newman. Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality. Physical Review E, 64(1): 2001.

[6] Z. Katona, P. P. Zubcsek, and M. Sarvary. Network Effects and Personal Influences: Diffusion of an Online Social Network. Journal of Marketing Research, 48(3): 425-443, 2011.

[7] D. H. Qiu, H. Li, and Y. Li. Identify Node Role and Track Node Evolution in Temporal Networks. In: Proceedings of International Conference of Soft Computing and Pattern Recognition (SoCPaR 2011), 129-134, 2011.

[8] C. Tantipathananandh, T. Y. Berger-Wolf, and D. Kempe. A Framework For Community Identification in Dynamic Social Networks. In: Proceedings of 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2007), 717-726, 2007.

[9] L. Tang, H. Liu, J. Zhang, and Z. Nazeri. Community Evolution in Dynamic Multi-mode Networks. In: Proceedings of 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), 677-685, 2008.

[10] L. Page. PageRank. Bring Order to the Web. Stanford Digtial Libraries Working Paper, 1997.

[11] W. Jiang, and G. Wang. SWTrust: Generating Trusted Graph for Trust Evaluation in Online Social Networks. In: Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011), 320-327, 2011.

[12] H. Liu, J. J. Salerno, and M. J. Young. Social Computing, Behavioral Modeling, and Prediction. US, Karsten Steinhaeuser, Nitesh V. Chawla, Springer, 168-175, 2008.

[13] G. Qin, and L. Gaoa. Spectral Clustering for Detecting Protein Complexes in Protein-Protein Interaction (PPI) Networks. Mathematical and Computer Modelling (Elsevier), 52: 2066-2074, 2010.

[14] P. W. Schmitz, and T. Troger. The (Sub-)Optimality of the Majority Rule. Games and Eco-nomic Behavior, Forthcoming, 2011.

[15] R. Merton. The Matthon Effect in Science. Science, New Series, 159: 56-63, 1968.

[16] J. Bai, and P. Wang. Conditional Markov Chain and Its Application in Economic Time Series Analysis. Journal of Applied Econometrics, 26(5): 715-734, 2011.

[17] A. S. Adeyefa, and M. K. Luhandjula. Multiobjective Stochastic Linear Programming: An Overview. American Journal of Operations Research, 203-213, 2011.

[18] I. Adan, R. D. Foley, and D. R. McDonald. Exact Asymptotics for the Stationary Distribution of a Markov Chain: A Production Model. Queueing Systems, 62: 311-343, 2009.

[19] J. Gehrke, P. Ginsparg, and J. M. Kleinberg. Overview of the 2003 KDD Cup. SIGKDD Explorations, 5(2): 149-151, 2003.

[20] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In: Proceedings of Internet Measurement Conference (IMC 2007), 29-42, 2007.

[21] W. Chen, Y. Wang, and S. Yang. Efficient Influence Maximization in Social Networks. In: Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2009), 199-207, 2009.