Information Sciences 379 (2017) 146–159

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Social influence modeling using information theory in mobile social networks

Sancheng Peng[a,b], Aimin Yang[a,*], Lihong Cao[c], Shui Yu[d], Dongqing Xie[e]

[a] School of Informatics, Guangdong University of Foreign Studies, Guangzhou, Guangdong Province 510420, PR China
[b] Laboratory of Language Engineering and Computing, Guangdong University of Foreign Studies, Guangzhou, Guangdong Province 510420, PR China
[c] School of English and Education, Guangdong University of Foreign Studies, Guangzhou, Guangdong Province 510420, PR China
[d] School of Information Technology, Deakin University, 221 Burwood HWY, Burwood VIC 3125, Australia
[e] School of Computer and Education Software, Guangzhou University, Guangzhou, Guangdong Province, 510006 PR China

ARTICLE INFO

Article history:
Received 15 November 2015
Revised 6 August 2016
Accepted 8 August 2016
Available online 12 August 2016

Keywords:
Mobile social networks
Social influence
Information entropy
social relationship graph
Direct influence
Indirect influence

ABSTRACT

Social influence analysis has become one of the most important technologies in modern information and service industries. Thus, how to measure social influence of one user on other users in a mobile social network is also becoming increasingly important. It is helpful to identify the influential users in mobile social networks, and also helpful to provide important insights into the design of social platforms and applications. However, social influence modeling is an open and challenging issue, and most evaluation models are focused on online social networks, but fail to characterize indirect influence. In this paper, we present a mechanism to quantitatively measure social influence in mobile social networks. We exploit the graph theory to construct a social relationship graph that establishes a solid foundation for the basic understandings of social influence. We present an evaluation model to measure both direct and indirect influence based on the social relationship graph, by introducing friend entropy and interaction frequency entropy to describe the complexity and uncertainty of social influence. Based on the epidemic model, we design an algorithm to characterize propagation dynamics process of social influence, and to evaluate the performance of our solution by using a customized program on the basis of a real-world SMS/MMS-based communication data set. The real world numerical simulations and analysis show that the proposed influence evaluation strategies can characterize the social influence of mobile social networks effectively and efficiently.

1. Introduction

Mobile social networks [6,19] are the networks where individuals with similar interests converse and connect with one another through their mobile phones and/or tablets. It is used widely as an important communication media with exponential growth. Especially, mobile communication and mobile web technologies have brought revolutionary changes to our daily lives. In the last decade, lots of online social networks, such as Facebook, Twitter, and LinkedIn, emerged and

* Corresponding author.
E-mail addresses: psc346@aliyun.com (S. Peng), amyang18@163.com (A. Yang).

http://dx.doi.org/10.1016/j.ins.2016.08.023
0020-0255/

tightly connected web users all over the world. People can directly engage in these networks, build their own friendship networks, and share their opinions, insights, information, experiences, and perspectives with each other. In addition, they can discover and spread information by using various formats, such as words, pictures, audios, and videos. Much like web-based social networks, mobile social networks occur in virtual communities. A current trend for social networking websites, such as Facebook, is to create mobile apps to provide their users instant and real-time access services through their devices.

Social influence [5] refers to the case when individuals change their behaviors under the influence of others. The strength of social influence depends on the relation among individuals, network distances, timing, characteristics of networks and individuals, etc. Viral marketing [10], online recommendation or advertising [13], healthcare community [37], expert finding [34], and other applications benefit from social influence by qualitatively and quantitatively measuring the influence of individuals on others. Social influence analysis is a highly utilizable technology, which is attracting a large number of researchers.

Modeling and analysis on social influence are becoming an important research field of social networks. By analyzing the influencing mode among users and the spreading mode of influence, many advantages can be obtained, such as being helpful to understand social behaviors of people from the angle of sociology; helpful to provide a theoretical basis for making public decision and public opinion guidance; and helpful to promote communication and dissemination of political, economic and cultural activities, as well as in other fields. Thus, social influence analysis in social networks has important social significance and application value.

Social influence analysis [24,32] is pervasive all over our society. The high-level goal in studying social influence analysis is to answer questions related to social influence, such as “Who can be influenced?”, “Who can influence whom?”, and “Who are the most influential users in a specific social network?”. The main problems of social influence analysis is how to quantify the influence of each user, and how to identify the most influential users in social networks. It has great potential to help us understand the ways in which information, experiences, ideas, and innovations propagate across social networks. Analyzing social networks can provide new insights into how people interact with and influence each other, and into why their ideas and opinions on different subjects can propagate in social networks.

As a natural consequence of the confusion in specific applications for measuring social influence, many solutions have been proposed. Some schemes on evaluation of social influence have been topic-oblivious [10,17,20,29,31,37,38,40,41]. In these works, social influence was measured either via the relative authority of individuals in their social networks, or via the degree of information diffusion with the social networks. Some schemes on evaluation of social influence have been topic-based [4,8,9,12,16,30,35,36,39]. In these works, social influence was measured by counting how much information related to a topic may be propagated in the networks. In addition, some schemes have been based on pairwise influence [1–3,11,13,15,18,21,26,28,33], which is defined based on social ties and interactions between users.

Although a variety of evaluation models of social influence are available, it is still not well understood what fundamental rules the evaluation models for social influence must follow. Without a good answer to this question, the design of evaluation models for social influence is still at the infant stage. Thus, we present a novel model to evaluate social influence of individuals based on entropy. In our proposed information theoretic framework, social influence is a measure of uncertainty, and their strength can be measured by entropy. We construct a social relationship graph by exploiting the graph theory to gain basic understanding of social influence. Based on the social relationship graph, we present a model to measure direct influence and indirect influence, by introducing friend entropy and interaction frequency entropy to describe the complexity and uncertainty of social influence. In addition, based on the epidemic model, we design an algorithm to characterize propagation dynamics process of social influence. Our goal is to design a general framework to characterize the social influence of each individual via a given information entropy. Our contributions are summarized as follows:

- We analyze the characteristics of the actual short message service/multimedia message service (SMS/MMS) communication data set from people’s daily lives for social interactions, and then construct a social relationship graph based on the methods of social network analysis. This graph is built to reveal the connections of social interaction and the spreading of SMS/MMS.
- We propose an evaluation model on social influence by using information entropy to reveal the relationship between social interactions and the strength of social influence. In this model, we measure direct influence and indirect influence, by integrating the friend entropy to evaluate the impact on social influence from the number of friends, and the interaction frequency entropy to evaluate the impact on social influence from the number of interactions.
- We design an algorithm to characterize propagation dynamics process of social influence by introducing epidemic model (i.e. susceptible-infectious model). In addition, we provide the performance evaluation for our solution by using a customized program on the basis of a real-world SMS/MMS-based communication data set. Through extensive numerical simulations and analysis, we confirm that our strategies can characterize the social influence of mobile social networks efficiently and effectively.

The remainder of this paper is organized as follows: In Section 2, we provide a survey of the related work, and formulate the problem of this paper in Section 3. In Section 4, we provide the framework of social influence analysis, and provide an overview of information entropy in Section 5. In Section 6, we present a model for measuring social influence based on information entropy, and describe the modeling on social influence propagation in Section 7. In Section 8, we conduct the performance evaluation, and conclude this paper and discuss future work in.

2. Related work

In this section, we investigate related work in three dimensions. The first dimension is the topic-oblivious influence evaluation model; the second is related to the topic-based influence evaluation model; and the last is related to the pairwise-based influence evaluation model.

(1) Topic-oblivious

Domingos and Richardson [10] investigated social influence in the customer network. They proposed a model to identify customer's influence between each other in the customer network, and built a probabilistic model to mine the spread of influence for viral marketing. Rodriguez et al. [29] proposed a model to infer the influence propagation given a set of observed user actions and associated timestamps to produce a network that best explains the observed actions times. Xiang et al. [40] developed an unsupervised model to estimate relationship strength from interaction activity (e.g., communication, tagging) and user similarity. They formulated a link-based latent variable model, along with a coordinate ascent optimization procedure for the inference. Tang et al. [37] proposed a model to compute the user influence by combining content-based and network-based approaches. They extended PageRank algorithm and proposed a UserRank algorithm to quantify user influence in a weighed social network.

Sathanur and Jandhyala [31] investigated the information-theoretic measure called transfer entropy as a measure of directed causal influence in online social interactions. Li and Gillet [17] evaluated the academic influence of scholars based on the scientific impact of their publications using three different measures, and investigated their social influence using network centrality metrics. Ye et al. [41] proposed a probabilistic generative model, namely social influenced selection (SIS), that explicitly quantifies and incorporates social influence from friends to a user, along with user preference and item content in the item selection process. Wang et al. [38] proposed a model, called dynamic social influence model, which simulates such social influencing processes that people dynamically change their attitudes when they communicate and exchange ideas with others.

(2) Topic-based

Tang et al. [35] proposed a Topical Affinity Propagation (TAP) method to model the topic-level social influence in social networks, and developed a parallel model learning algorithm based on the map-reduce programming model. Cha et al. [4] exploited three measures such as number of followers, number of mentions and number of retweets, to quantify a user's influence. They also compared the relevance of three measures in the form of correlation coefficient.

Weng et al. [39] presented an algorithm called Twitter-Rank based on PageRank algorithm to measure the influence of users in Twitter. TwitterRank measures the influence taking both the topical similarity between users and the link structure into account. Ding et al. [9] measured the influence of users using random walks on the multi-relational data (i.e. the retweet, the reply, the reintroduce, and the read) in Micro-blogging. Li et al. [16] proposed a probabilistic model to capture the dual effect of topic preference and to mine topic-level opinion influence in microblog. Sang and Xu [30] presented a multimodal topic-sensitive influence model, which enables simultaneous extraction of node topic distribution, topic-sensitive edge strength, and the topic space.

Tang et al. [36] studied a problem of conformity influence analysis in large social networks. They defined three major types of conformities to formulate the problem of conformity influence analysis. Herzig et al. [12] proposed an Author-Reader Influence (ARI) model to evaluate the influence of various users on others by applying a retrospective analysis from an ordinary reader's point of view. Cui et al. [8] presented a Hybrid Factor Non-Negative Matrix Factorization (HF-NMF) approach for modeling item-level social influence.

(3) Pairwise-based

Pal et al. [21] used topic-sensitive retweet information to quantify pairwise influence and to identify the topic authorities in Twitter. Anagnostopoulos et al. [1] exploited a statistical analysis method to identify and measure whether social influence is a source of correlation between the actions of individuals with social ties. Peng et al. [26] introduced two factors to evaluate influence of each node. One factor is intimacy degree (ID), which is used to reflect the closeness between users. The other factor is activity degree (AD), which is used to determine which node is more active. He et al. [11] introduced transfer entropy to identify peer influence in online social networks. Bakshy et al. [3] quantified pairwise influence in the form of an influence score. They also quantified the global influence of a user by summing all his or her influence scores, and constructed disjoint influence tree model with some features to predict the user's global influence. Huang et al. [13] proposed a framework to quantitatively measure individuals' social influence by examining the number of their followers and their sensitivity of discovering good items.

Aral and Walker [2] proposed a method by using vivo randomized experimentation to identify influence and susceptibility in networks. Kwak et al. [15] exploited three different measures of influence, such as number of followers, pagerank, and number of retweets, to measure influence. Su et al. [33] developed an algorithm based on the PageRank algorithm, called InfluentialRank (IR), which calculates the influence of nodes based on the following relationship of users, retweet behaviours, and users' interests. Phan et al. [28] presented the Topic-aware Community-level Physical Activity Propagation (TaCPP) model, to capture the social influences of messages in the YesiWell study. Li et al. [18] proposed a conductance eigenvector centrality (CEC) model to measure peer influence in social networks.

3. Problem formation

In this section, we introduce several related definitions, and then formally formulate the problem of evaluation modeling on social influence in mobile social networks.

The problem of social influence modeling has been open for a long time since it was proposed. It is because social influence is a relatively fuzzy concept and lacks of universally acknowledged definition. People are frequently confused by the concept and the measuring method of social influence. For different social networks, the meaning of social influence and the metrics used to measure may be quite different [7]. In electronic commerce network, the most influential individuals are the people who can recommend most customers to purchase specific product successfully. In microblog network, the most influential bloggers are these who attract more people to share and comment on. In SMS/MMS-based networks, the most influential customers are the active people who communicate with many friends frequently by sending the SMS/MMS messages in each period time. After analyzing the characteristics and the way of information diffusion in different social networks, we formally give the definitions of mobile social networks and social influence.

Definition 1 (Mobile social network)**.** Mobile social network is defined as a directed, weighted graph, $G(V, E, W)$, where set V of vertices corresponds to mobile terminals in a wireless network, set E of directed edges corresponds to the traffic flow among mobile terminals i to j, and set W of weight values corresponds to the total number of interactions made from i to j.

A mobile social network [22,25,26] is a social network extracted from a data set of mobile communication, where each cellular phone represents a node, and relations between any two cellular phones form the set of edges. With the rapid development and increasing popularity of mobile social networks, more and more interest has been made in leveraging information available from mobile social networks to promote the adoption of new products or services. For example, in a mobile social network, Bob usually gathers information from his friends, when he contemplates purchasing products or services. Bob also shares his opinions within a specific mobile social networks regarding to different products which he has recently purchased or he is familiar with.

Definition 2 (Direct Influence)**.** Given two individuals u and v in a mobile social network, who are directly connected in the network, u has the capability to change or impact the opinion of v in a direct way. $DI_u(t)$ is defined as the direct influence of user u on its one-hop friends.

Definition 3 (Indirect Influence)**.** Given two individuals u and v in a mobile social network, who are not directly connected in the network, u has the capability to change or impact the opinion of v in an indirect way. $II_{uv}(t)$ is defined as the indirect influence of user u on v.

Definition 4 (Global Influence)**.** Given a mobile social network, and u exerts the power over the whole network, $I_u(t)$ is defined as the global influence of u, which represents the global influential strength of u over the whole network.

Social influence is a relationship established between two entities for a specific action. In particular, one entity influences the other entity by performing actions. In this work, the first entity is called the *influencer*, and the second entity is called the *influencee*. We introduce the notations to describe a social influence relationship. The level of social influence can be measured by a continuous real number, referred to as the social influence value. Social influence can also be represented by uncertainty (i.e., strongest, stronger, strong, medium, weak, weaker, weakest).

Therefore, the problem of social influence modeling could be converted to the study of computing direct influence $DI_u(t)$, indirect influence $II_{uv}(t)$, and global influence $I_u(t)$, for any individual u in a given mobile social network $G(V, E, W)$. In addition, in the existing models for social influence evaluation, the uncertainty of social influence in mobile social networks has been largely ignored. Besides the direct influence, another interesting question is "Does the influence exist between users who are not connected?" That is, "Does a user have a certain indirect influence on his/her friends' in a mobile social network?" In this paper, we present a novel model to measure social influence, which includes direct influence and indirect influence, using information entropy to describe the complexity and uncertainty of social influence.

4. Framework of social influence analysis

The framework of social influence analysis in mobile social networks is shown in Fig. 1.

Data collection from mobile social networks: It is a very important basis of social influence analysis. With the availability of mobile social networking big data generated by advertising, instant messaging, mobile communication, the collection of raw data from online sources (e.g., Facebook) and offline sources (e.g., call data) is much easier.

Data preprocessing: In order to improve the performance and the convenience of processing, we need to remove the irrelevant information on the social influence analysis. In order to protect the privacy of users, we also need to filter out the sensitive information from the collected data sets.

Selection of evaluation metrics: It is very important to select a set of evaluation metrics to describe accurately the characteristics of each user, as these metrics are helpful to measure social influence of each user and to find easily the most influential top-k nodes. In the existing work, the common evaluation metrics include centrality of each node, interaction frequency among odes, etc.

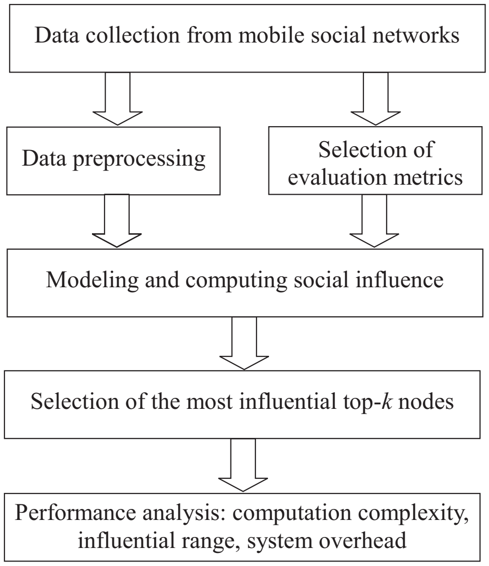

Fig. 1. Framework of social influence analysis.

Modeling and computing social influence: According to the selected evaluation metrics, evaluation models and computing equations on social influence are provided to a specific mobile social network. Thus, social influence of each user can be measured, by integrating the related computing equations into the collected real-world data sets from mobile social networks.

Selection of the most influential top-*k* nodes: The influence maximization algorithm is designed to find the most influential top-k nodes. Most of the existing algorithms are the improved algorithms based on the greedy algorithm.

Performance analysis: Simulation is made to validate performance (e.g. computational complexity, influential range, system overhead) of the proposed algorithms based on a specific propagation model. For an influence maximization algorithm, it has low computational complexity and low system overhead, and the selected most influential top-k nodes has a large influential range.

5. Preliminary of information entropy

Information entropy was founded by Claude Elwood Shannon, who is an American mathematician, electronic engineer and cryptographer, in his work “A Mathematical Theory of Communication” in 1948. It is a concept from information theory. It tells how much information there is in an event. In general, the more uncertain or random the event is, the more information it will contain. It has applications in many areas, including lossless data compression, statistical inference, cryptography, and recently in other disciplines as biology, physics or machine learning, and so on.

According to Shannon’s theory, if a random variable X represents a set of possible events x_i whose probabilities of occurrence are p_i, $i = 1, \ldots, n$, then a measure $H(X)$ of the uncertainty of the outcome of an event given such distribution of probabilities should have the following three properties [14]:

- $H(x_i)$ should be continuous in p_i;
- If all probabilities are equal, which means that $p_i = 1/n$, then $H(X)$ should be a monotonically increasing function of n (if there are more choices of events, then the uncertainty about one outcome should increase);
- If a choice is broken down into other choices, with probabilities c_j, $j = 1, \ldots, k$, then $H(X) = \sum_{j=1}^{k} c_j H_k$, where H_k is the value of the function $H(X)$ for each choice.

Thus, Shannon proved that the only function that satisfies all three properties is given by:

$$H(X) = H(x_1, x_2, \cdots, x_n) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i), \tag{1}$$

where the sum is over all states (Shannon’s definition had a constant k multiplied by it, which has been removed here). The base 2 for the logarithm is chosen so that the measure is given in terms of bits of information. For example, a device with two positions (like a flip-flop circuit) can store one bit of information. The number of possible states for n such devices would then be 2^n, and $\log_2 2^n = n$, meaning that n such devices can store n bits of information, as should be expected. This

Table 1
The number of interactions between two cellular phone users in a week.

Between two smartphones	*The number of interactions*	*Between two smartphones*	*The number of interactions*
$A \to B$	3	$F \to I$	7
$A \to C$	6	$G \to B$	7
$A \to D$	9	$G \to D$	8
$B \to A$	6	$G \to F$	8
$B \to D$	15	$G \to H$	21
$B \to F$	4	$G \to I$	0
$B \to G$	8	$G \to J$	5
$C \to A$	16	$H \to D$	3
$C \to D$	7	$H \to E$	11
$C \to E$	11	$H \to G$	13
$D \to A$	4	$H \to J$	7
$D \to B$	6	$H \to K$	3
$D \to C$	7	$I \to F$	6
$D \to E$	3	$I \to G$	7
$D \to G$	4	$I \to J$	12
$D \to H$	17	$J \to G$	9
$E \to C$	4	$J \to H$	22
$E \to D$	8	$J \to I$	11
$E \to H$	8	$J \to K$	6
$E \to K$	0	$K \to E$	4
$F \to B$	2	$K \to H$	0
$F \to G$	1	$K \to J$	2

definition bears a lot of resemblance to Gibbs' entropy, but is more general, as it can be applied to any system that carries information.

6. Social influence modeling using information entropy

6.1. Constructing social relationship graph

We model a mobile social network using a directed, weighted graph, $G(V, E, W)$, where set V of vertices corresponds to the cellular phones in cellular networks, set E of directed edges corresponds to the traffic flow among cellular phones i to j, and set W of weight values corresponds to the total number of SMS/MMS messages sent from cellular phone i to j in a given time period. The degree of vertex i, d_i, is the number of cellular phones that it communicates with. The amount of messages initiated from i to j is denoted by C_{ij}.

To construct social relationship graphs, we collected a big data set from one of the largest cellular networks in China. The data set of 0.4 million users in this network exchanged about 20 million SMS/MMS messages over a three-week period in October 2012. To protect the privacy of users, the content of the messages is removed, and the uniqueness of the identifiers of the phone numbers involved are replaced by pseudocodes.

In our previous work [23], we have analyzed the characteristics of the complex network for the real-world data set, and have found that the degrees of the nodes for the probability density distribution of in-degree and out-degree obey a power law distribution, which indicates that the mobile social network is a scale-free network.

From the real-world data set, we can extract a smartphone social network. The network is huge and complex. In order to explain the idea of a smartphone social network, we take eleven users from the data set and use them as an example. The data of this sample social network is listed in Table 1.

According to Table 1, we treat each smartphone as a vertex, so a weighted social relationship graph can be obtained and is shown in Fig. 2. From Fig. 2, we find that social-interactions based on SMS/MMS are in our everyday lives. If only the one-way behavior is presented, even if W is large, it is difficult to show that the relationship between these two users are very close. For example, if a smartphone user always sends advertising information to other users, it cannot show that their relationships are close. Thus, we take the smaller of the weight values of the two directed edges between two nodes to measure the strength of their social relationship. That is, the social graph weight between node i and j at time t, denoted by $W_{ij}(t)$, is given by

$$W_{ij}(t) = \min\{C_{ij}(t), C_{ji}(t)\}, \tag{2}$$

where $C_{ij}(t)$ denotes the number of SMS/MMS messages sent from node i to j at time t.

Thus, according to Eq. (2), we can obtain the effective interactions between two nodes in a week, as shown in Table 2.

In addition, the transition result of Fig. 2 is shown in Fig. 3. The behaviors of SMS/MMS-based social interactions between two smartphone users are characterized accurately by changing one directed weighted graph $G(V, E, W)$ into another undirected weighted graph $G'(V, E, W)$.

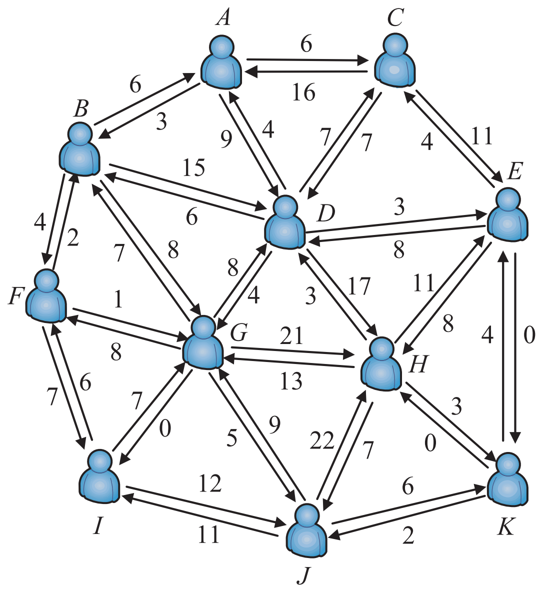

Fig. 2. A weighted social relationships graph (directed) based on the total number of SMS/MMS messages between two nodes in a week.

Table 2
The number of effective interactions between two cellular phone users in a week.

Between two smartphones	*The number of effective interactions*	*Between two smartphones*	*The number of effective interactions*
A, *B*	3	*E*, *H*	8
A, *C*	6	*E*, *K*	0
A, *D*	4	*F*, *G*	1
B, *D*	6	*F*, *I*	6
B, *F*	2	*G*, *H*	13
B, *G*	7	*G*, *I*	0
C, *D*	7	*G*, *J*	5
C, *E*	4	*H*, *J*	7
D, *E*	3	*H*, *K*	0
D, *G*	4	*I*, *J*	11
D, *H*	3	*J*, *K*	2

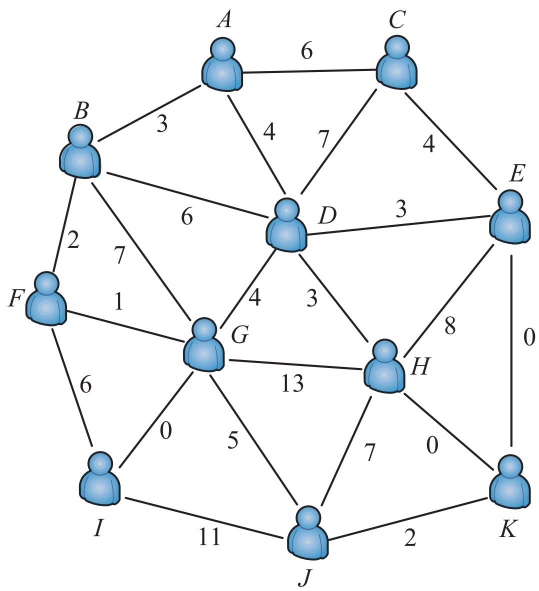

Fig. 3. A weighted social relationships graph (undirected) based on the number of effective interactions between two nodes in a week.

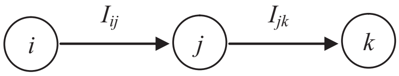

Fig. 4. Single path.

6.2. Computing on direct influence

1) Entropy computation among friend nodes

First, we use a function $f_{ij}(t)$ to characterize the friendship between node i and j at time t.

$$f_{ij}(t) = \begin{cases} 1, & W_{ij}(t) > 0 \\ 0, & W_{ij}(t) = 0 \end{cases} \tag{3}$$

Let $N_i(t)$ be the number of one-hop friend nodes of node i at time t. In general, the size of $N_i(t)$ is an important factor to measure the influence of a node in a mobile social network. Now, let N be the total number of nodes in a mobile social network. Thus, $N_i(t)$ is described as follows.

$$N_i(t) = \sum_{j=1}^{N} f_{ij}(t). \tag{4}$$

Thus, the entropy of friend nodes $I_i^f(t)$ for node i is described as follows.

$$I_i^f(t) = -\sum\nolimits_{i=1}^{N_i(t)} \frac{1}{N_i(t)} \log_{10} \frac{1}{N_i(t)} \tag{5}$$

2) Entropy computation on interaction frequency among friend nodes

In general, the size of $C_{ij}(t)$ is also an important factor for entropy computation on interaction frequency among friend nodes. Thus, the entropy of interaction frequency $I_i^c(t)$ for node i is described as follows.

$$I_i^c(t) = -\sum\nolimits_{j=1}^{N_i(t)} \frac{C_{ij}(t)}{\sum_{k=1}^{N_i(t)} C_{ik}(t)} \log_{10} \frac{C_{ij}(t)}{\sum_{k=1}^{N_i(t)} C_{ik}(t)} \tag{6}$$

Thus, the total direct influence of i on its one-hop friend nodes is described as follows.

$$DI_i(t) = \alpha I_i^f(t) + \beta I_i^c(t), \tag{7}$$

where α and β denote the weight of $I_i^f(t)$ and $I_i^c(t)$, respectively, and $\alpha + \beta = 1$.

6.3. Computing on indirect influence

Let $N_{ik}(t)$ represent the common friend nodes between i and k. It is described as follows.

$$N_{ik}(t) = N_i(t) \cap N_k(t), \tag{8}$$

where $N_i(t)$ and $N_k(t)$ represent one-hop friend nodes of i and k, respectively.

1) If $N_{ik}(t) = 0$, there is no path from node i to k.

2) If $N_{ik}(t) = 1$, there is a path from node i to k, which is shown in Fig. 4. That is, i has one two-hop friend node k, or k has one two-hop friend node i.

Thus, the indirect influence of node i on k is described as follows.

$$II_{ik}(t) = I_{ij} \times I_{jk} = DI_i(t) \times DI_j(t) \tag{9}$$

3) If $N_{ik}(t) = 3$, there are three paths from node i to k, which is shown in Fig. 5.

Thus, the indirect influence of node i on k is described as follows.

$$\begin{aligned} II_{ik}(t) &= \frac{I_{ij}(t) \times I_{jk}(t) + I_{im}(t) \times I_{mk}(t) + I_{in}(t) \times I_{nk}(t)}{3} \\ &= \frac{DI_i(t) \times DI_j(t) + DI_i(t) \times DI_m(t) + DI_i(t) \times DI_n(t)}{3} \end{aligned} \tag{10}$$

According to the above analysis, the indirect influence is described as follows.

$$II_{ik}(t) = \sum\nolimits_{k=1}^{N_{ik}(t)} DI_i(t) * DI_j(t) / N_{ik}(t) \tag{11}$$

Let $M_i(t)$ be the number of two-hop friend nodes of i at time t. Thus, the total indirect influence $II_i(t)$ of i is described as follows.

$$II_i(t) = \frac{\sum_{k=1}^{M_i(t)} II_{ik}(t)}{M_i(t)} \tag{12}$$

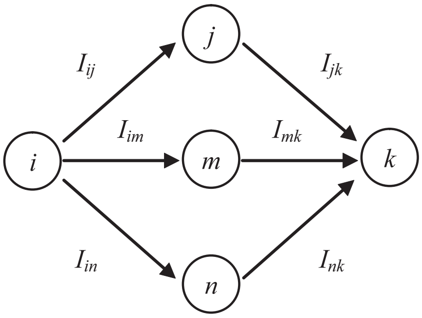

Fig. 5. Multiple path.

Table 3
The number of friend nodes for each node.

Terminal	*The number of friend nodes*
A	3
B	4
C	3
D	6
E	3
F	3
G	5
H	4
I	2
J	4
K	1

Let us take Fig. 2 as an example, the indirect influence of node A is given by: $II_A(t) = (DI_A(t) \times DI_B(t) + DI_A(t) \times DI_D(t) + (DI_A(t) \times DI_B(t) + DI_A(t) \times DI_D(t))/2 + (DI_A(t) \times DI_C(t) + DI_A(t) \times DI_D(t))/2)/4$

6.4. Total influence of node

According to the above analysis, the total influence $I_i(t)$ of i is described as follows.

$$I_i(t) = \omega_1 DI_i(t) + \omega_2 II_i(t), \tag{13}$$

where ω_1 and ω_2 denote the weight of the total direct influence $DI_i(t)$ and the total indirect influence $II_i(t)$, respectively, and $\omega_1 + \omega_2 = 1$.

Algorithm 1 shows the complete computing process of influence for all nodes.

6.5. An example for measuring social influence

According to Fig. 3, the number of friend nodes for each node is listed in Table 3.

Algorithm 1 Influence computing algorithm for all nodes.

Input: A social network $G(V, E, W)$, a set of evaluation measure;
Output: A social network $G'(V, E, W)$ with each node i being assigned an influence value $I_i(t)$;

for i = 1 to E **do**
  Constructs friend relationship network;
end for
for i = 1 to N **do**
  Computes the entropy of friend nodes $I_i^f(t)$ using Eq. (5);
  Computes the entropy of interaction frequency $I_i^c(t)$ using Eq. (6);
  Computes the total direct influence $DI_i(t)$ of i on its one-hop friend nodes using Eq. (7);
  Computes the total indirect influence $II_i(t)$ using Eq. (12);
  Computes the total influence $I_i(t)$ using Eq. (13);
end for

Table 4
Value of social influence for each node.

Terminal	*Direct influence*	*Indirect influence*	*Total influence*
A	0.4544	0.2887	0.3881
B	0.5711	0.3509	0.4830
C	0.4632	0.2954	0.3961
D	0.7265	0.4480	0.6151
E	0.4657	0.2961	0.3979
F	0.3998	0.2322	0.3327
G	0.6616	0.3964	0.5555
H	0.5723	0.3720	0.4922
I	0.2862	0.1390	0.2273
J	0.5715	0.3321	0.4757
K	0	0	0

Table 5
The results of sorting.

Node	*No.*
D	1
G	2
H	3
B	4
J	5
E	6
C	7
A	8
F	9
I	10
K	11

In the light of Eq. (7), we set particular values for the following parameters: α=0.4, and β =0.6, the total direct influence of each node on its one-hop friend nodes is listed in Table 4. Moreover, in the light of Eq. (12), the total indirect influence of each node on its two-hop friend nodes is listed in Table 4.

We set particular values for two parameters: ω_1=0.6 and ω_2 =0.4. According to Eq. (13), the total social influence of each nodes is listed in Table 4. On the basis of the value of social influence for each node, the results of the sorting are listed in Table 5.

7. Modeling on social influence propagation

It is well known that the classical influence diffusion models in social networks include linear threshold model (LTM), independent cascade model (ICM), and weighted cascade model (WCM). However, influence diffusion model can also be seen as a specific case of the traditional epidemic models. In our previous work presented in Ref. [27], we have surveyed the epidemic models in detail. In this paper, we introduce SI (susceptible-infectious) model to characterize the propagation dynamics process of social influence. In SI epidemic model, we suppose a susceptible individual i (i.e., influencee), after successful contact with an infectious individual j (i.e., influencer), becomes infected. This represents j influences i. Let T be the transmission threshold through which a node i transforms from state S to state I.

Integrating the characteristics of social influence propagation with the social relationship of users, we design a propagation model to characterize the process of SMS/MMS-based social influence propagation. The specific steps of this model are described as follows: 1) to evaluate the social influence of each node by using Algorithm 1; 2) to find the most influential top k nodes by sorting influence value of nodes with the heap sort strategy; and 3) to characterize state transition of nodes by using Algorithm 2.

Analysis for time complexity: Let d denote the average number of one-hop friend nodes, n denote the total number of nodes in the network, $n = |V|$, k denote the number of the most influential nodes, and m denote the total number of edges, $m = |E|$. Thus, the time complexity of Algorithm 1 and Algorithm 2 is $O(m + nd)$, $O(m + nd + n\log(k))$, respectively.

8. Performance evaluation

To validate the effectiveness of the proposed model, we conduct experiments by using the message records collected by one of the largest cellular networks in China. In addition, we design and develop a C++ simulator to implement our proposed mechanism. Due to the huge scale of the real-world data set, we take 5114 users by using graph connectivity determination algorithm for our experiments, rather than include all the users.

Algorithm 2 State transition algorithm for all nodes.

Input: A social network $G(V, E, W)$, a set of most influential k nodes;
Output: Influence spread at time t in a social network for a given set of most influential k nodes;

Network initialization. Compute social influence for each node i using Algorithm 1;
Node state initialization. The most influential k nodes are selected with the heap sort algorithm, and their states are set to be state I, and the states of other nodes are set to be state S;
The information of its/theirs friends can be collected by counting the messaging records;
Node i is accessed at time t, thus
while $i \le N$ **do**
  if (Its state is I) **then**
    Its neighbor nodes are accessed;
    while $j \le N_i$ **do**
      if (The state of its friend node j is S) and ($I_j(t)$ is not smaller than T) **then**
        Node j changes its state from S to I;
      else
        Node j remains in its previous state;
      end if
    end while
  end if
end while
t equals to t plus Δt;

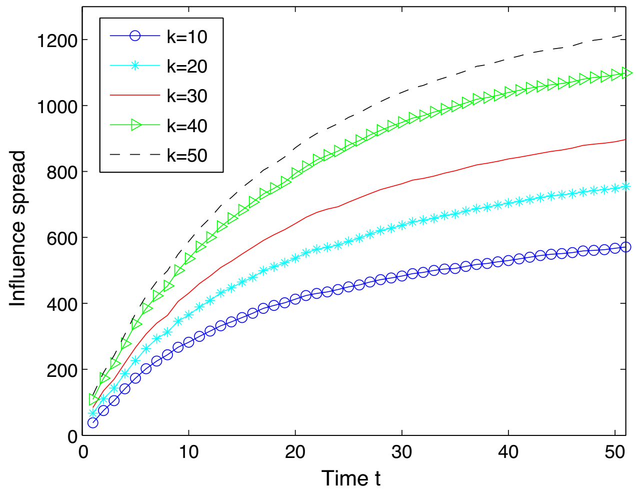

Fig. 6. A comparison of influence spread of the entropy-based social influence evaluation model with different k influential nodes.

In the experiments, we compare our proposed algorithm with the random algorithm and node degree algorithm for spreading of social influence. These algorithms are listed as follows:

Random: As a baseline comparison, it simply selects k nodes in network as seed nodes.

Degree: As a comparison, it simply selects k nodes with the largest degrees in network as seed nodes.

Entropy-based: The entropy-based influence evaluation model (Algorithm 2) selects k nodes with the most influential in network as seed nodes.

The influence diffusion is a metric to measure how many users can be influenced by the most influential k specific users (or called seed nodes). To test the influence spread, we use the SI model to propagate social influence. To obtain the influence spread of each algorithm, for each seed set, we run the simulation of the worm propagation model on the networks 100 times and take the average number of the influence spread. In addition, to obtain the influence spread of each model, we first select top $k = (10, 20, 30, 40, 50)$ influential nodes as seeds.

Fig. 6 shows the influence spread of the entropy-based social influence evaluation model with different k at time t. As can be seen from the results, as the value of k increases, the number of influence spread increases. It is because the more the most influential nodes there are, the more nodes can be influenced.

Fig. 7 shows the influence spread of different algorithms with different k influential nodes. From the results, it can be seen that the random algorithm as the baseline performs very badly, and the degree-based algorithm is better but is still significantly worse than the entropy-based algorithms. This shows that relying only on structural properties of the network

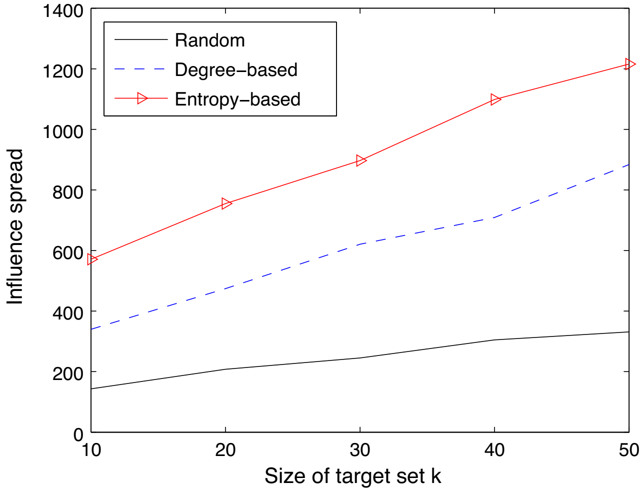

Fig. 7. A comparison of influence spread of different models with different $k = (10, 20, 30, 40, 50)$ influential nodes.

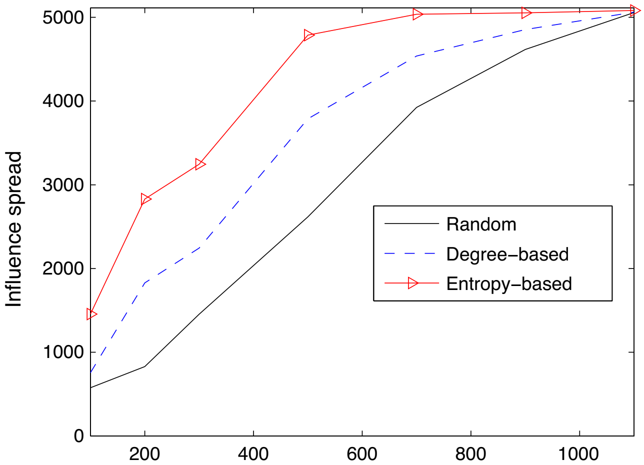

Fig. 8. A comparison of influence spread of different models with different $k = (100, 200, 300, 500, 700, 900, 1100)$ influential nodes.

can not obtain better results. In addition, the influence spread increases slowly, as k changes from 1 to 5. Then as the value of k increases, the number of influence spread increases quickly. This is because only the top 1 to 5 nodes are influential nodes and the succeeding nodes do not contribute to increasing the influence spread.

From Fig. 7, with the increase of the value of k, we find that the three curves may not come close. To explore when k equals to which value, these curves will come closer and closer. Thus, we select top $k = (100, 200, 300, 500, 700, 900, 1100)$ influential nodes as seed sets.

From the results in Fig. 8, it can be seen that the influence spread of the entropy-based algorithm increases quickly from $k = 100$ to 500, the degree-based algorithm increases quickly from $k = 200$ to 800, and the influence spread of the random algorithm increases quickly from $k = 300$ to 1000. In general, the influence spread of the entropy-based algorithm increases more quickly than that of the random algorithm and that of the degree-based algorithm. However, as $k = 1100$, the influence spread of these three algorithms is coming closer and closer to 5114.

9. Conclusion and future work

In this paper, we presented a framework to quantify social influence in mobile social networks. The social influence of users was measured by analyzing the SMS/MMS-based communication behaviors among individuals. In addition, we revealed and characterized the social relations among mobile users through the analysis on the entropy of friend nodes and the entropy of interaction frequency. The extensive analytical results demonstrate that the influence spread of our proposed method outperforms that of the random method and that of the degree-based method.

As for our further work, we will focus on characterizing the impact of casual relationship on social influence, and distinguishing positive influence and negative influence. In addition, how to quickly and effectively identify influential users in mobile social network is also worthy of further study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61379041 and 61300195, and the key lab of Guangzhou Mobile Internet Security and Fault Tolerance under grant No. 2014SY000022.

References

[1] A. Anagnostopoulos, R. Kumar, M. Mahdian, Influence and correlation in social networks, in: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2008), Vancouver, 2008.
[2] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, Science 337 (6092) (2012) 337–341.
[3] E. Bakshy, J.M. Hofman, W.A. Mason, D.J. Watts, Everyone's an influencer: Quantifying influence on twitter, in: Proceedings of the 4th ACM international conference on web search and data mining (WSDM 2011), Hong Kong, China, 2011, pp. 65–74.
[4] M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummad, Measuring user influence on twitter: The million follower fallacy, in: Proceedings of the 4th international AAAI conference on weblogs and social media (ICWSM 2010), Washington, DC, USA, 2010.
[5] M. Chen, S. Mao, Y. Liu, Big data: a survey, Mobile Netw. Appl. 19 (2014) 171–209.
[6] S. Chen, G. Wang, W. Jia, k-fuzzytrust effient trust computation for large-scale mobile social networks using a fuzzy implicit social graph, Inf. Sci. 318 (2015) 123–143.
[7] W. Chen, S. Cheng, X. He, F. Jiang, Influencerank: An e cient social influence measurement for millions of users in microblog, in: Proceedings of the second international conference on cloud and green computing (CGC2012), Xiangtan, Hunan, China, 2012, pp. 463–470.
[8] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, L. Sun, Who should share what?: Item-level social influence prediction for users and postsranking, in: Proceedings of the 34th international ACM conference on research and development in information retrieval (SIGIR 2011), Beijing, China, 2011, pp. 185–194.
[9] Z. Ding, Y. Jia, B. Zhou, Y. Han, Mining topical influencers based on the multi-relational network in micro-blogging sites, China Commun. 10 (1) (2013) 93–104.
[10] P. Domingos, M. Riehardson, Mining the network value of customers, in: Proceedings of the 7th ACM conference on knowledge discovery and data mining, New York, USA, 2001, pp. 57–66.
[11] S. He, X. Zheng, D. Zeng, K. Cui, Z. Zhang, C. Luo, Identifying peer influence in online social networks using transfer entropy, in: Proceedings of pacific asia workshop intelligence and security informatics (PAISI 2013), Beijing, China, 2013, pp. 47–61.
[12] J. Herzig, Y. Mass, H. Roitman, An author-reader influence model for detecting topic-based influencers in social media, in: Proceedings of the 25th ACM conference on hypertext and social media (HT 2014), Santiago, Chile, 2014, pp. 46–55.
[13] J. Huang, X. Cheng, H. Shen, T. Zhou, X. Jin, Exploring social influence via posterior effect of word-of-mouth recommendations, in: Proceedings of the fifth ACM international conference on web search and data mining (WSDM 2012), Seattle, Washington, USA, 2012, pp. 573–582.
[14] L.S. Jr, Structure of a global network of financial companies based on transfer entropy, Entropy 16 (2014) 4443–4482.
[15] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: Proceedings of the 19th international conference on world wide web (WWW 2010), raleigh, NC, USA, 2010, pp. 591–600.
[16] D. Li, X. Shuai, G. Sun, J. Yang, Y. Ding, Z. Luo, Mining topic-level opinion influence in microblog, in: Proceedings of the 21st ACM international conference on information and knowledge management (CIKM 2012), Maui, USA, 2012, pp. 1562–1566.
[17] N. Li, D. Gillet, Identifying influential scholars in academic social media platforms, in: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2013), 2013, pp. 608–614.
[18] X. Li, Y. Liu, Y. Jiang, X. Liu, Identifying social influence in complex networks: A novel conductance eigenvector centrality model, Neuro comput. (2016), doi:10.1016/j.neucom.2015.11.123.
[19] L. Lin, L. Xu, S. Zhou, Y. Xiang, Trustworthiness-hypercube-based reliable communication in mobile social networks, Inf. Sci. (2016), doi:10.1016/j.ins.2016.05.048.
[20] Q. Liu, B. Xiang, E. Chen, H. Xiong, F. Tang, J.X. Yu, Influence maximization over large-scale social networks: A bounded linear approach, in: Proceedings of the 23rd ACM international conference on conference on information and knowledge management (CIKM 2014), Shanghai, China, 2014, pp. 171–180.
[21] A. Pal, C. Scott, Identifying topical authorities in microblogs, in: Proceedings of the fourth ACM international conference on web search and data mining (WSDM 2011), Hong Kong, 2011, pp. 45–54.
[22] S. Peng, J. Li, A. Yang, Entropy-based social influence evaluation in mobile social networks, in: Proceedings of the 15th international conference on algorithms and architectures for parallel processing (ICA3PP 2015), zhangjiajie, china, 2015, pp. 637–647.
[23] S. Peng, M. Wu, G. Wang, S. Yu, Containing smartphone worm propagation with an influence maximization algorithm, Comput. Netw. 74 (2014) 103–113.
[24] S. Peng, G. Wang, D. Xie, Social influence analysis in social networking big data: Opportunities and challenges, IEEE netw. (2015).
[25] S. Peng, G. Wang, S. Yu, Modeling malware propagation in smartphone social networks, Proceedings of the 12th IEEE international conference on trust, security and privacy in computing and communications (trustcom 2013), Australia, 196–201, 2013.
[26] S. Peng, G. Wang, S. Yu, Mining mechanism of top-k influential nodes based on voting algorithm in mobile social networks, in: Proceedings of the 11th IEEE/IFIP international conference on embedded and ubiquitous computing (EUC 2013), Zhangjiajie, China, 2013, pp. 2194–2199.
[27] S. Peng, S. Yu, A. Yang, Smartphone malware and its propagation modeling: A survey, IEEE Commun. Surv. Tut. 16 (2) (2014) 925–941.
[28] N. Phan, J. Ebrahimi, D. Kil, B. Piniewski, D. Dou, Topic-aware physical activity propagation in a health social network, IEE Intell. Syst. (2016) 5–14.
[29] M.G. Rodriguez, J. Leskovec, A. Krause, Inferring networks of diffusion and influence, in: Proceeding of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (KDD 2010), 2010, pp. 1019–1028.
[30] J. Sang, C. Xu, Social influence analysis and application on multimedia sharing websites, ACM transactions on multimedia computing, Commun. Appl. 9 (1s) (2013) 1–24.
[31] A.V. Sathanur, V. Jandhyala, An activity-based information-theoretic annotation of social graphs, in: Proceedings of the 2014 ACM conference on web science (websci 2014), Bloomington, USA, 2014, pp. 187–191.
[32] A. Saulwick, K. Trentelman, Towards a formal semantics of social influence, Knowledge-Based Syst. 71 (2014) 52–60.
[33] C. Su, Y. Du, X. Guan, C. Wu, Maximizing topic propagation driven by multiple user nodes in micro-blogging, in: Proceedings of the 38th annual IEEE conference on local computer networks, 2013, pp. 751–754.
[34] J. Sun, W. Xu, J. Ma, J. Sun, Leverage RAF to find domain experts on research social networks services: A big data analytics methodology with mapreduce framework, Int. J. Product. Econ. 165 (2015) 185–193.
[35] J. Tang, J. Sun, C. Wang, Z. Yang, Social influence analysis in large-scale networks, in: Proceeding of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD 2009), New York, NY, USA, 2009, pp. 807–816.
[36] J. Tang, S. Wu, J. Sun, Confluence: conformity influence in large social networks, in: Proceeding of the 19th ACM SIGKDD international conference on knowledge discovery and data mining (KDD 2013), Chicago, Illinois, USA, 2013, pp. 347–355.
[37] X. Tang, C.C. Yang, Ranking user influence in healthcare social media, ACM Trans. Intell. Sys. Technol. 3 (4) (2012) 1–21 Article No. 73.

[38] Z. Wang, R. Shinkuma, T. Takahashi, Dynamic social influence modeling from perspective of gray-scale mixing process, in: Proceedings of the eighth international conference on mobile computing and ubiquitous networking (ICMU 2015), Hakodate City, Hokkaido, Japan, 2015, pp. 1–6.
[39] J. Weng, E.-P. Lim, J. Jiang, Q. He, Twitterrank: finding topic-sensitive influential twitterers, in: Proceedings of the third ACM international conference on web search and data mining (WSDM 2010), New York, USA, 2010, pp. 261–270.
[40] R. Xiang, J. Neville, M. Rogati, Modeling relationship strength in online social networks, in: Proceedings of the 19th international world wide web conference (WWW 2010), Raleigh, NC, USA, 2010, pp. 981–990.
[41] M. Ye, X. Liu, W.-C. Lee., Exploring social influence for recommendation: a generative model approach, in: Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval, Portland, Oregon, USA, 2012, pp. 671–680.