Information Sciences 396 (2017) 114–126

Contents lists available at ScienceDirect

Information Sciences

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

A user behavior influence model of social hotspot under implicit link

Yunpeng Xiao[a,*], Na Li[a], Ming Xu[b], Yanbing Liu[a]

[a] Chongqing Engineering laboratory of Internet and Information Security, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[b] Research Institute of Information Technology, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 8 April 2016
Revised 17 November 2016
Accepted 15 February 2017
Available online 20 February 2017

Keywords:
Social network
Hotspot topic
Influence analysis
Implicit link
Cloud model

ABSTRACT

In social networks, user behavior is affected by complex dynamic factors. Here, we investigate the internal and external factors that drive users to participate in social hotspot s. By analyzing user behavior, we discover the differences between driving factors and quantify their driving strength. First, four factors that influence the user's behavior are proposed, including explicit links (*E*), implicit links (*I*), personal interest (*P*), and a random factor (*R*). In particular, based on a cloud model, an implicit link creation method is designed. This method can quantify the driving strength of the implicit relation between users, and avoid the multiple attribute weighting defects in subjective and objective aspects. Next, considering the maximum likelihood estimation theory, a user behavior influence model (*EIPR*) of a hotspot topic is proposed to measure the causes of user behavior behind the social hotspots. Experimental results show that the model can be used to find different dynamic factors of user behavior in social hot topics. Among these external factors, the implicit link plays an significantly important role in driving user behavior.

1. Introduction

What drives human behavior? This question has inspired scientists for hundreds of years. Today, in the internet and big data era, online network systems such as BBS and social networks have gradually become important platforms for investigating user behavior. In this study, we try to mine the influence factors that drive user behavior on the background of social hotspots.

The online social network is increasingly important today for social relations maintenance and user behavior diffusion in human society. In fact, hotspot topics are promoted by several factors in a social network. Discovering these factors has a profound impact on social development, social services and management, and social security. At present, the study about user behavior influence is divided into two aspects: internal influence and external influence. For the internal influence factors, Barabasi et al. [2,22] first discovered the long tail phenomenon of user behavior. And they proposed a user behavior model based on task priority to analyze the influence factors that drive user behavior. Subsequently, a variety of user behavior influence model are proposed by scholars for different application scenarios. The study of internal factors that affect user behavior is gradually clear through above research.

* Corresponding author.
E-mail address: xiaoyp@cqupt.edu.cn (Y. Xiao).

http://dx.doi.org/10.1016/j.ins.2017.02.035
0020-0255/

In addition, the external factors that influence user behavior are studied by scholars from the aspects of user behavior interaction, the neighbor node influence, common interests and behavior game etc. However, the basis of these studies is link between the social users, here we name this kind of link as explicit link. Compared to previous studies, new features of online social networking are emerging, which bring more challenges to existing methods, mainly in the following: 1) complex network structure. Group level interaction is not a simple accumulation of individual behaviors. Especially for group behavior relating to hot topics, explicit relations between online users cannot reflect real group network structure characteristics, and there are several implicit and complex relations between the groups both online and offline. 2) the group behavior dynamics are influenced by complex factors. Researches on offline individual user behavior are already challenging. Moreover, studying online group behavior requires knowledge on multiple subjects, such as computer science, sociology, and psychology. Consequently, a number of unknown network-group behavior rules require further exploration.

In this study, combined with the internal and external factors, the user participation behavior of social hot topics is deeply analyzed. While discussing the interaction between users, we especially try to mine the effect of implicit links as the external factor that affects user behavior. The influence factors that drive user behavior–explicit link factor, implicit link factor, personal interest factor, and random factor are extracted for modeling and analysis. Our goal is to analyze the differences between internal and external factors affecting different users, to quantify the effect of these factors, and to mine the role of the implicit link in driving user behavior.

Our contribution can be summarized as follows:

- To analyze the behavior of users to participate in hotspot topics, we propose four factors: the explicit link factor, implicit link factor, personal interest factor, and random factor. To analyze and quantify these factors, the *EIPR* influence model is proposed. It is found that different users are affected by different influence factors and the influence strength is also different, as demonstrated through model analysis.
- A scheme for creating implicit link based on the cloud model is proposed. In this scheme, we give a more clear definition for implicit link. Meanwhile, the defects in the traditional weighting methods with regard to subjective and the objecting weighting and the advantage of the cloud model in overcoming this problem are considered. The implicit user relation and strength of the relation are being quantified.
- The proposed method cannot only mine the dynamic factors that drive user behavior, but also make a contrast analysis between the effects of the explicit and implicit factors, which can promote topic evolution. This work fully embodies the influence of the tacit role. In addition, it helps in managing and controlling the public sentiment and providing support to locate the online water army.

The rest of this paper is organized as follows. In Section 2, we introduce the related work. In Section 3, we formulate the problem and give the necessary definitions. In Section 4, we describe the model algorithm to analyze user behavior. In Section 5, we present and analyze the experimental results of the model.

Finally, we conclude this work in Section 6.

2. Related work

In a social network, user behavior is influenced by many factors. After the task priority model of user behavior influence was proposed by Barabasi [2,22], the scholars put forward different influence factors and their behavior model for different application scenarios, such as browsing the web [34], watching online movies [39], personal interests [20] and trust agents [18]. At present, for individual influence discovery, researchers aim to find the difference in the individuals and define the opinion of leaders in a group by modeling [23]. To measure the network structure based individual influence, early methods were mainly based on the degree centrality [37], closeness [36] and betweenness [14]. In addition, combined with individual influence discovery and random walk [12,24], some representative algorithms were proposed, such as HITS [11], and PageRank [21]. Wang et al. [27] studied the evolution of user behavior over time, and found a behavior influence analysis of low-literate users of a viral speech based telephone service. Individual influence is based on user behavior. Thus, some researchers determined the relation between users through analyzing the information diffusion and user behavior [13,15], including copying, replying, retweeting and so on. To study individual influence based on topic, the analysis was focused on topic evolution and some topic model is proposed [5,9]. The abovementioned research on the influence of individual behavior provides the academic accumulation for the following work.

For the external factors that affect user behavior, most scholars pursue research on the basis of the network structure, mainly in the following aspects: 1) Network structure based analysis of influence strength. In order to measure the strength of individual influence on Twitter, the following network and the forwarding network was analyzed [4]. Zhu, et al. [10]proposed a user behavior model that users affected by neighbor nodes in terrorist incidents. 2) Calculation of influence strength based network structure. In order to find out the information dissemination path of messages, the relations between edge structure and information pathway in the network was analyzed [7,30]. 3) User behavior based measurement of influence strength. Wu et al. [31,32] quantified the factors that affect information dissemination and topic evolution by analyzing the user interaction behavior in social networks. In the calculation of the influence strength based on topics, Tang et al. [25] studied the role of different groups in topic evolution. Meanwhile, they quantified the influence strength and predicted the evolution of topics. In addition, the implicit relation of users can be analyzed by combining the topic distribution, information diffusion and influence [8,19].

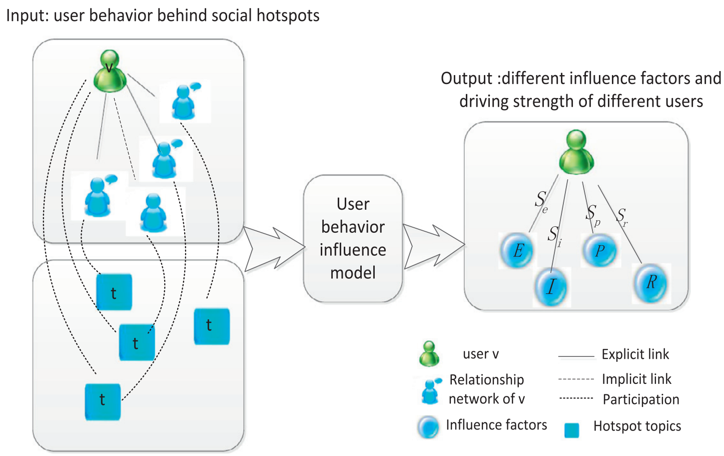


Fig. 1. Problem overview.

The above mentioned research were conducted based on the explicit link between users. Relation strength is considered when measuring the influence between users. Besides, the implicit relation between users is usually evaluated by user similarity. Bakshy et al. [1,3] showed the reason for information diffusion and sharing in social media by analyzing the strong and weak relation between users. They found that weak connections probably play a more important role in online information diffusion. Wang et al. [28] proposed the model of EAGR to analyze the weighting problem between adjacent points. Wang et al. [26] combined influence and node similarity measurement to model and analyze the most influential nodes in the network.

However, in the actual network behavior, it is possible that there is an implicit relationship among users–implicit link. The implicit link means implicit relationship, between users, which does not own nor generate explicit link in the future. However, we can mine the implicit relationship from the behind of user behavior. For example, a topic hype team members may never create a link, but they have been mutual cooperation. Therefore, this implicit link cannot be calculated with traditional similarity method or link prediction. Dong et al. [6] mentioned the implicit link when they studied link prediction based on a heterogeneous network. However, the implicit link they mentioned is just a type of relationship mapping in heterogeneous networks, which is completely different from the concept of implicit link in this study. Alternately, Xiao et al. [33] have proposed the concept of implicit link, but they defined the implicit relationship only by simple statistical methods. In this paper, the implicit link will be further quantified based on the previous research. Furthermore, the influence factors that drive user behavior will be comprehensively further analyzed on the background of social network hot topics.

3. Problem definition

In the present study, we attempt to mine the factors that drive user v to participate in hotspots by analyzing the user's attribute and relation network, as shown in Fig. 1. Let $G(v, E)$ be the explicit relation network of user v, that is, the following network. E is the edges connected to user v and $U(E)$ is the set of users who have an explicit relation with v. Then, we define $A=(T_m, v, t)$ as the released topic data of user v, where T_m is the topic data that v released in time t. $B=\{(T_P, v\&U(E), t)\}$ is the topic data of user v and $U(E)$ joined in a different time. Specifically, $\left(T_p, v\&U(E), t\right)$ means that user v and $U(E)$ participated in topics T_p in time t.

3.1. Input

The input in this paper consists of: 1) the explicit relation network of user v, $G(v, E)$; 2) the topics that user v generated, $A=(T_m, v, t)$; and 3) the topics that user v and E participated in, $B=\{(T_P, v\&U(E), t)\}$.

The purpose of this paper is to make personalized analysis for each user, as well as to mine and quantify the influence factors that affect different users. In addition, we analyze the role of the implicit link in driving users to participate in hotspot topics. Taking advantage of the cloud model in weighting multiple factor weighting, a method to create the implicit link based on the cloud model is proposed.

3.2. Related definitions

3.2.1. Driving factors

We extract the influence factors explicit links factor (E), implicit link factor (I), personal interest factor (P) and random factor (R) from the data on personal attributes, the relational network, and historical behavior. The specific relationship between the factors and user v is shown in Fig. 2.

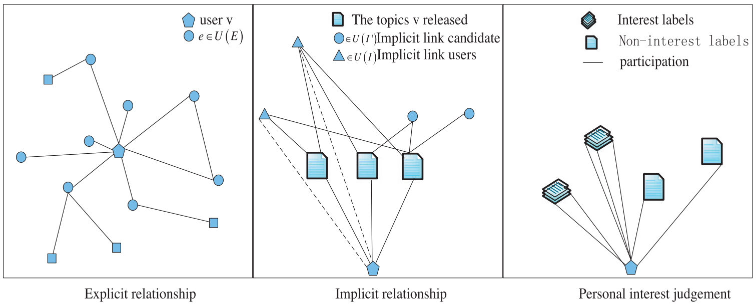


Fig. 2. The relation between user v and the influence factor.

Definition 1 (The explicit link driving factor α_e).

$$\alpha_e = J_e(T_k, t_v, t_{U(E)})/N_{T(E)} \tag{1}$$

$$J_e(T_k, t_v, t_{U(E)}) = \begin{cases} 1 & t_v \le t_{U(E)} \\ 0 & t_v > t_{U(E)} \end{cases} \tag{2}$$

$J_e(T_k,\ t_v,\ t_{U(E)})$ represents whether the time $t_{U(E)}$ is earlier than time t_v. $t_{U(E)}$ is the earliest time when users in set $U(E)$ participate d in topic T_k. T_k refers to the topic that user v participated in. t_v represents the time when user v participate in topic T_k . $N_{T(E)}$ is the number of the topics users in set $U(E)$ participated in.

Definition 2 (The implicit link driving factor α_i).

$$\alpha_i = \left(T_k, t_v, t_{U(I)}\right)/N_{T(I)} \tag{3}$$

$$J_i\left(T_k, t_v, t_{U(I)}\right) = \begin{cases} 1 & t_v \le t_{U(I)} \\ 0 & t_v > t_{U(I)} \end{cases} \tag{4}$$

$J_i(T_k,\ t_v,\ t_{U(I)})$ represents whether the time $t_{U(I)}$ is earlier than time t_v. $t_{U(I)}$ is the earliest time when users in set $U(I)$ participate d in topic T_k. T_k refers to the topic that user v participated in. t_v represents the time when user v participate d in topic T_k. $N_{T(I)}$ is the number of topics users in set $U(I)$ participated in.

Definition 3 (The personal interest driving factor α_p).

$$\alpha_p = J_p(T_k)/N_{T(p)} \tag{5}$$

$$J_p(T_k) = \begin{cases} 1 & N_l \ge \ln N_T \\ 0 & N_l < \ln N_T \end{cases} \tag{6}$$

$$N_{T(p)} = g\omega \sum_{L} 1 \tag{7}$$

$J_p(T_k)$ represents whether the label l of topic T_k belongs to L where L is the interest label set of user v. N_l is the number of topics with label l that user v participated in, and N_T is the total number of topics user v participated in. Let $N_{T(p)}$ represent the number of topics with interest label. g is the increment of topics number under every interest label, and ω is the mean topics duration time.

Definition 4 (The random driving factor α_r).

$$\alpha_r = 1/N_{T(R)} \tag{8}$$

$N_{T(R)}$ is the number of topics that user v can participate in randomly.

3.2.2. Definition about cloud model

Due to the deficiencies of the probability theory and fuzzy mathematics in dealing with uncertain data, Li et al. [16] proposed a cloud model that illustrates the transformation of qualitative and quantitative data. The cloud model can not only represent a qualitative concept with a quantitative value, but also describe quantitative value with qualitative language. In addition, the application of the model in various fields has achieved significant results [29,35].

Definition 5 (Cloud and cloud droplet)**.** We assume U is a quantitative domain, and C is a qualitative value. The degree of membership $u(u \in [0, 1])$ is a random number with a stable tendency which represents the qualitative relation between x and C, where $x \in U$. The distribution of the membership degree $u(u \in [0, 1])$ in the domain U is called the membership cloud, cloud for short. Thus, a cloud is a mapping from domain U to interval [0,1]. $x \in U$ is called a cloud droplet.

The integral form of the cloud is determined by its digital features, including expectations, entropy, and hyper entropy. These features reflect the quantitative characteristics of the qualitative concepts, and they are defined as follows respectively.

$$Ex = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{9}$$

$$En = \sqrt{\frac{\pi}{2}} * \frac{1}{N} * \sum_{n=1}^{N} |X_n - Ex| \tag{10}$$

$$He = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (X_n - Ex)^2 - En^2} \tag{11}$$

Definition 6 (Cloud generator)**.** Assume that U is a quantitative domain represented by precise value and C is the oriented concept of U. If the quantitative value x satisfies $x \in U$, x is a random realization of C with a qualitative concept. If $x \sim N\left(Ex, En'^2\right)$, where $E_n \sim e^{-\frac{(x-Ex)^2}{2\left(En'\right)^2}}$ and the certainty degree μ of x for C satisfied is $\mu = e^{-\frac{(x-Ex)^2}{2\left(En'\right)^2}}$, the distribution of x in the quantitative domain U is called the normal cloud.

Definition 7 (Cloud similarity $S_c(C_1, C_2)$)**.** Cloud similarity measurement is a method to measure the similarity of clouds by calculating their droplet distances [17,38]. If two clouds C_1 and C_2 generated a fixed number of cloud droplets through a normal cloud generator, we regard the mean distance of all corresponding cloud droplets in C_1 and C_2 as the similarity index value between C_1 and C_2.

3.3. Problem

Problem. Given $G(v, E)$, $G(v, I)$ and $B = \{(T_P, v\&U(E), t)\}$, we solve the following problems.

1. How can the implicit link (I) be defined and measured ? How can we mine the implicit relation among users and give quantitative analysis using the concept of implicit link.
2. How can we model and analyze the difference of factors in driving different users to participate in topics, and how can we measure the driving strength.

4. Model

4.1. Framework

The model framework is shown in Fig. 3. First, we analyze the data on user attributes, relational network, and historical behavior in the social network to extract the influence factors. Second, according to the influence factors, the four driving factors α_e , α_i, α_p and α_r are defined. Finally, the *EIPR* influence model is proposed to analyze the difference between the influence factors and to quantify their driving strength. Meanwhile, the role of the implicit link in driving user behavior is further analyzed.

4.2. Establishing the implicit link based on the cloud model

To solve the first problem, we extract the influence factors from the personal attributes, relational network, and historical behavior. Considering the particularity of the implicit link factor, an implicit link establishment method is given.

Step 1. Improve the definition of the implicit link.

Implicit link: It is a game relation between users generated from different opposites, or a n implicit relation generated due to a common purpose or interest. Besides, there is no explicit relation between users. For example, user A and B have different positions on topic T. They argue with each other so that the two users connect and interact, though they don't have any explicit relationship.

Step 2. Candidate users extraction for the implicit link.

Users who participated in the topics released by v are extracted as a set $U(I')$. $U(I')$ is the set of users who are likely to create an implicit link with v and $i' \in U(I')$. The existence of an implicit link is determined by the similarity between v and i'. In addition, $i' \in U(I')$, $i' \notin U(E)$.

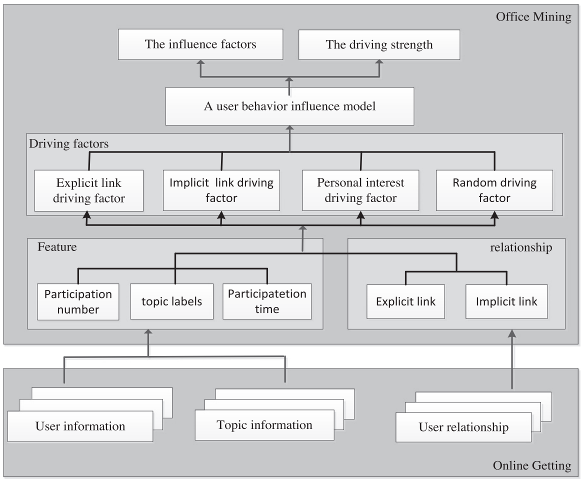


Fig. 3. Model framework.

Step 3. User behavior cloud generation.

Topics and labels have many-to-one relation. If the user v participated in the topics of the same label with i', v and i' have the same label. Different candidate users i' have different quantities of the same label with user v. We assume that v and i' have two common labels l_1 and l_2 to analyze the relation between them. According to the quantity of topics that v and i' participated in, the numerical features of l_1, l_2 can be separately calculated. $v_{l_1} : \left\{Ex_{l_1}, En_{l_1}, He_{l_1}\right\}$ and $v_{l_2} : \left\{Ex_{l_2}, En_{l_2}, He_{l_2}\right\}$ are the digital features of user v in l_1 and l_2. Similarly, $i'_{l_1} : \left\{Ex_{l_1}, En_{l_1}, He_{l_1}\right\}$ and $i'_{l_2} : \left\{Ex_{l_2}, En_{l_2}, He_{l_2}\right\}$ are the digital features of user i'. The digital features can be calculated as follows:

$$Ex_l = \frac{1}{M} * \sum_{m=1}^{M} x_m \tag{12}$$

$$En_l = \sqrt{\frac{\pi}{2}} * \frac{1}{M} * \sum_{m=1}^{M} |x_m - Ex_l| \tag{13}$$

$$He_l = \sqrt{\frac{1}{M}\sum_{m=1}^{M} (x_m - Ex_l)^2 - {En_l}^2} \tag{14}$$

M is the summation of the time slices divided by the month. l means a same label of user v and i'. In addition, to ignore the impact of activity on the user's participation in topics, we normalize the number of topics that users participated in under the same label to get x_m. Then according to the digital features of v, we obtain one thousand cloud droplets by cloud generator. The same method can be used to generate the could droplets of i'. x, y are the normal random numbers we get by the cloud generator. The user behavior cloud is formed form the spatial distribution of cloud droplets.

$$\overrightarrow{\alpha_v(j)} = \left(x_{vj}, y_{vj}, \mu_v(j)\right) \quad j \in [1, 1000] \tag{15}$$

$$\overrightarrow{\alpha_{i'}(j)} = \left(x_{i'j}, y_{i'j}, \mu_{i'}(j)\right) \quad j \in [1, 1000] \tag{16}$$

Step 4. Similarity calculation for C_v and C'_i.

In order to make a comprehensive evaluation of user behavior, we calculate the behavior cloud similarity of C_v and $C_{i'}$ using the cloud similarity algorithm. As user behavior cloud similarity of user behavior, the calculation of $S_c(v, i')$ is as follows:

$$S_c(v, i') = 1 - \frac{1}{N} * \sum_{i=1}^{N} |y_v(m) - y_{i'}(n)| \tag{17}$$

$$[m, n] = \min\left\{\sqrt{(x_{vm} - x_{i'n})^2 + (y_{vm} - y_{i'n})^2}\right\} m, n \in [1, 1000] \tag{18}$$

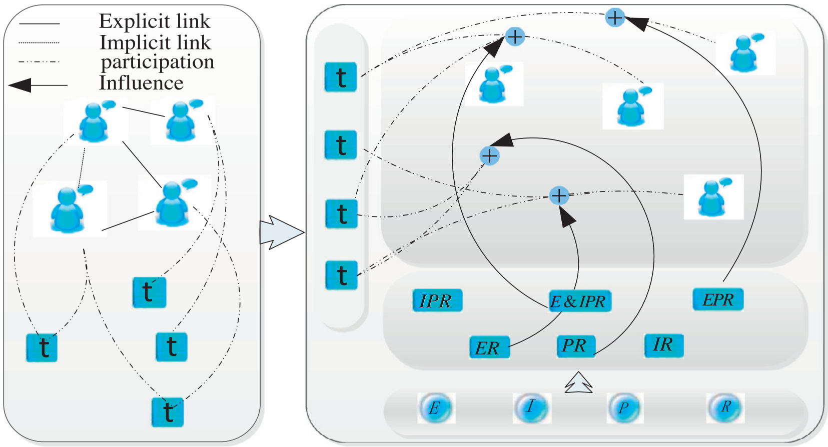


Fig. 4. Model detail.

$|y_v(m) - y_{i'}(n)|$ is the difference of the membership degree between two corresponding cloud droplets in C_v and $C_{i'}$. The corresponding droplets mean they should satisfy min $\left\{\sqrt{(x_{vm} - x_{i'n})^2 + (y_{vm} - y_{i'n})^2}\right\}$ and m, $n \in [1, 1000]$. And N denotes the quantity of cloud droplets in cloud C_v or $C_{i'}$.

Step 5. Implicit links establishment.

Similarity between user v and i' is calculated, $i' \in U(I')$. Then, we rank the similarity to obtain the set $U(I)$ of users whose similarity with v are in the top $0.5N_{I'}$. The implicit link between the users v and i' is created to establish the implicit relation network $G(v, I)$ of user v, where $i \in U(I)$. The user similarity is calculated as follows:

$$S(v, i') = \frac{L_v \cap L_{i'}}{L_v \cup L_{i'}} * S_c(v, i') \tag{19}$$

$N_{I'}$ is the number of user in the set of $i' \in U(I')$. $\frac{L_v \cap L_{i'}}{L_v \cup L_{i'}}$ represents the Jaccard coefficient of the labels between users v and i'. $S(v, i')$ is the similarity of the users v and i'.

4.3. Influence model

To solve the second problem, the influence model *EIPR* is proposed based on the maximum likelihood theory. We convert the user behavior to the relevant probability event and estimate its parameters through *EIPR*. The model detail are shown in Fig. 4.

First, we assume that the behavior of user v to participate in hotspot topics is driven by independent factor. Under this assumption, we analyze the behavior of user v in an independent drive situation, including the patterns of *ER*, *IR* and *PR*. The *ER*, *IR* and *PR* pattern s assume that the behavior of v is only driven by the explicit link factor, the implicit link factor or the personal interest factor. Moreover, we assume that the behavior of user v to participate in hotspot topics is driven by a combination of these factors. Under this assumption, we analyze the behavior of v in combined drive situations, including the patterns of *EPR*, *IPR* and *E& IPR*. In the *EPR* pattern, the behavior of user v to participate hotspot topics is driven by the explicit link factor and personal interest factor together. In the *IPR* pattern, the behavior of v is driven by the implicit link factor and personal interest factor. In the *E& IPR* pattern, the behavior of user v is driven by the explicit link factor, the implicit factor and the personal interest factor. Finally, we obtain the most suitable pattern for user v and the driving strength of the factors in that pattern to analyze the role of the implicit link futher.

1. Independent drive

Assuming that user v is driven only by the explicit link factor, the analysis with the pattern *ER* is as follow s:

$$\begin{aligned} L_e(S_e) &= \prod_{T(v)} (S_e * \alpha_e + (1 - S_e) * \alpha_r) \\ &= \prod_{J_e(T_k, t_v, t_{U(E)})=1} \left(\frac{S_e}{N_{T(E)}} + \frac{1 - S_e}{N_{T(R)}}\right) \prod_{J_e(T_k, t_v, t_{U(E)})=0} \frac{1 - S_e}{N_{T(R)}} \end{aligned} \tag{20}$$

The estimated parameter S_e represents the driving strength of the the explicit link factor. $T(v)$ refers to the set of topics that user v participated in.

$$LnL_e(S_e) = \sum_{J_e(T_k,t_v,t_{U(E)})=1} \ln\left(\frac{S_e}{N_{T(E)}} + \frac{1-S_e}{N_{T(R)}}\right) + \sum_{J_e(T_k,t_v,t_{U(E)})=0} \ln\left(\frac{1-S_e}{N_{T(R)}}\right) \tag{21}$$

According to the theory of maximum likelihood estimation, S_e is the driving strength of the explicit link factor when $LnL_e(S_e)$ reachers the maximum value $\max nL_e(S_e)$ in the *ER* pattern. Similarly, the driving strength s of the implicit link factor and the personal interest factor can be obtained when $LnL_i(S_i)$ and $LnL_p(S_p)$ obtain their respective maximum values $\max LnL_i(S_i)$ and $\max LnL_p(S_p)$ in the patterns *IR* and *PR*.

2. Combined drive

Assuming that user v is driven by the explicit link factor and personal interest factor, the analysis with the pattern *EPR* is as follows:

$$L_{ep}(S_e, S_p) = \prod_{T(v)} (S_e * \alpha_e + S_p * \alpha_p + (1 - S_e - S_p) * \alpha_r) \tag{22}$$

$$\begin{aligned} LnL_{ep}(S_e, S_p) = & \sum_{\substack{J_e(T_k,t_v,t_{U(E)})=1\\ J_p(T_k)=1}} Ln\left(\frac{S_e}{N_{T(E)}} + \frac{S_p}{N_{T(P)}} + \frac{(1-S_e-S_p)}{N_{T(R)}}\right) \\ & + \sum_{\substack{J_e(T_k,t_v,t_{U(E)})=1\\ J_p(t_k)=0}} Ln\left(\frac{S_e}{N_{T(E)}} + \frac{1-S_e-S_p}{N_{T(R)}}\right) \\ & + \sum_{\substack{J_e(T_k,t_v,t_{U(E)})=0\\ J_p(t_k)=1}} Ln\left(\frac{S_p}{N_{T(P)}} + \frac{1-S_e-S_p}{N_{T(R)}}\right) \\ & + \sum_{\substack{J_e(T_k,t_v,t_{U(E)})=0\\ J_p(t_k)=0}} Ln\left(\frac{1-S_e-S_p}{N_{T(R)}}\right) \end{aligned} \tag{23}$$

S_e and S_p are the driving strength of the explicit link factor and the implicit link factor, when $LnL_{ep}(S_e, S_p)$ get the maximum value $\max LnL_{ep}(S_e, S_p)$ in the *EPR* pattern. Similarly, the driving strength of the implicit link factor and the personal interest factor can be obtained when $LnL_{ip}(S_i, S_p)$ reaches the maximum value $\max LnL_{ip}(S_i, S_p)$ in the *IPR* pattern.

In addition, we analyze the *E& IPR* pattern that the explicit link factor, the implicit link and the personal interest factor combined.

$$L_{e\&ip}(S_{e\&i}, S_p) = \prod_{T(v)} \left(S_{e\&i} * \frac{J_{ei}(T_k, t_v, t_{EI})}{N_{T(E)\cup T(I)}} + S_p * \frac{J_P(T_k)}{N_{T(P)}} + \frac{1-S_{e\&i}-S_p}{N_{T(R)}}\right) \tag{24}$$

$N_{T(E)\cup T(I)}$ is the number of topics that users in the set of and participated in. $S_{e\&i}$ and S_p are the driving strength when $LnL_{e\&ip}(S_{e\&i}, S_p)$ reaches the maximum value $\max LnL_{e\&ip}(S_{e\&i}, S_p)$ in the *E& IPR* pattern. Otherwise, it is difficult to obtain the optimal solution analytically. We numerically explore the values of the two parameters in the unit square to maximize the log-likelihood.

4.4. Model algorithm

In this paper, we model user behavior to determine the factors that drive users to participate in hotspot topics and quantify their driving strength. Meanwhile, the role of the implicit link factor in driving user behavior is further explored. The model algorithm is as follows.

input :

1. Network : $G(v, E)$;

2. User behavior in different time: $A = (T_m, v, t)$, $B = \{(T_P, v\&U(E), t)\}$

output:

1. The driving factors $\{E, I, P, R\}$ of user v;

2. The driving strength of that factors;

Initialize $N_{T(E)}$, $N_{T(I)}$, $N_{T(P)}$, $N_{T(R)}$

for *each topic* t_k *of user* v *participated in* **do**

  //calculate the driving factors α_e, α_i, α_p, α_r.

  if v *participated in* t_k *based* E *or* I *or* P **then**

    $J_{\mathrm{e}}(T_k, t_v, t_{U(E)}) = 1$, $J_{\mathrm{e}}(T_k, t_v, t_{U(I)}) = 1$, $J_p(T_k, t_v, t_P) = 1$

  end

  else

    $J_{\mathrm{e}}(T_k, t_v, t_{U(E)}) = 0$, $J_{\mathrm{e}}(T_k, t_v, t_{U(I)}) = 0$, $J_p(T_k, t_v, t_P) = 0$

  end

  obtain $\alpha_e = J_e(T_k, t_v, t_{U(E)})/N_{T(E)}$, $\alpha_i = J_i(T_k, t_v, t_{U(I)})/N_{T(I)}$,
  $\alpha_p = J_p(T_k)/N_{T(p)}$, $\alpha_r = 1/N_{T(R)}$

  //Independent drive.For example EP pattern
  $L_e(S_e)* = (S_e * \alpha_e + S_r * \alpha_r)$

  //Combined drive. For example EPR pattern
  $L_{ep}(S_e, S_p)* = (S_e * \alpha_e + S_p * \alpha_p + S_r * \alpha_r)$

end

//compute the parameters and the maximum likelihood values

$$result = \max \left\{ \begin{array}{l} \max LnL_e, \max LnL_i, \max LnL_p, \\ \max LnL_{ep}, \max LnL_{ip}, \max LnL_{e\&ip} \end{array} \right\}$$

In the model algorithm, the parameters $N_{T(E)}$, $N_{T(I)}$, $N_{T(P)}$ and $N_{T(R)}$ are initialized to calculate the driving factors α_e, α_i, α_p, and α_r. Then, we determine the most suitable pattern for user v by analyzing the six patterns in the independent drive and combined drive situations. According to the most suitable pattern, we can capture the driving factor that drives user to participate hotspot topics and its driving strength. In addition, the complexity of the algorithm is considered. The extraction of the driving factors $T_{extract} = O(n)$; the analysis of the patterns $T_{patterns} = O(n)$; and the calculation of the result $T_{result} = O(1)$. Based on the above analysis, the total time complexity of the model algorithm is $T = T_{extract} + T_{patterns} + T_{result} = O(n)$.

5. Experimental results

5.1. Data sets

In this study, the experimental data was collected from the social platform of *tianya* forum. The forum of *tianya* is one of the most popular social networking platforms in China. The total number of topic labels in the forum was 482. The number of topics user v can participate in randomly reached 61,480,000. The increment of topics reached 100,000 every day. The average life cycle ω of topics is considering seven days. In order to carry out a comprehensive analysis of the influence factors, we collected the historical behavior and relation data of 1437 users in the past 12 months . Due to space constraints, use two users as examples to illustrate the results of our analysis. The related data of the two users are shown in Table 1.

User A: The first user is *xibeixiongying*. The related data (topics and relation) for him on December 1, 2014 and October 31, 2015 are used as the fundamental data for model. The number of related users is 11,272, and the number of topics is 52,030.

User B: The second user is *huachengxiaoliu*. The related data(topics and relation) for him on December 1, 2014 and October 31th 2015 are used as the fundamental data for model .The number of related users is up to 1732, and the number of topics is 663,553.

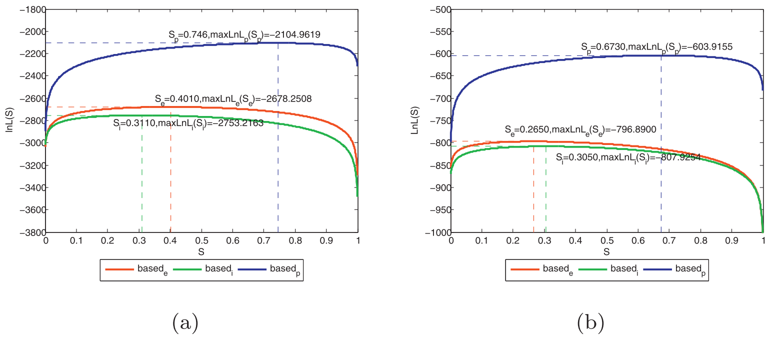


Fig. 5. (a).Independent drive of user A. (b). Independent drive of user B.

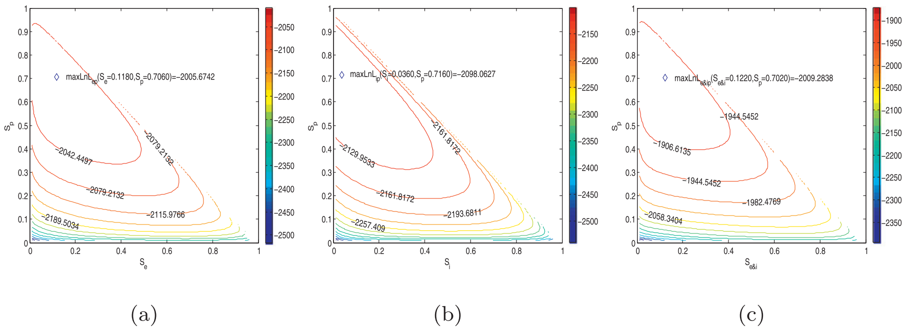


Fig. 6. (a) *EPR* for user A; (b) *IPR* for user A; (c) *E& IPR* for user A.

Table 1
Data sets of user A and B.

Data sets	Nodes	Explicit nodes	Implicitnodes	Topics	Topics (v)	Topics (explicit)	Topics (implicit)
User A	11,272	11,212	60	52,030	403	25,307	26,330
User B	1732	394	1337	663,553	208	34,184	629,369

Table 2
The driving strength of factors in six patterns of user A.

Situation	Model	Strength	maxlnL
Independent drive	ER	Se = 0.4010	−2678.2508
	IR	Si = 0.3110	−2753.2163
	PR	Sp = 0.7460	−2104.9619
Combined drive	EPR	Se = 0.1180, Sp = 0.7060	−2005.6420
	IPR	Si = 0.0360, Sp = 0.7160	−2098.0627
	E&IPR	$S_{e\&i} = 0.1220$, Sp = 0.7020	−2009.2838

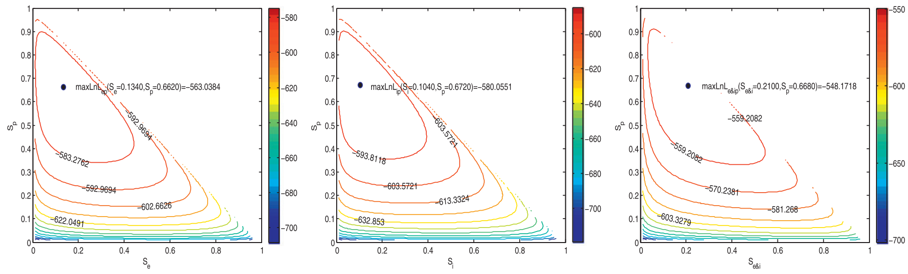
5.2. Driving performance analysis

By observing the relation between user activity and the frequency that each user participates in topics, it becomes evident that the amount of data increases with the activity. Consequently we analyze the users that the number of topics they participate in is up to 50 in the latest year.

First, we show the results of three patterns *ER, IR* and *PR* in the independent drive situation. Then, we illustrate the results for the three patterns in the combined drive situation. Finally, we determine the most suitable pattern for the user and obtain the driving strength of every factor in that patterns.

The experimental results for the two users in independent drive situation are shown in Fig. 5. The x-axis represents the driving strength, and the y-axis represents the maximum likelihood function in the different driving strength.

As shown in Fig. 5(a), user A has a higher maximum likelihood $\max \ln L(S)$ in pattern *PR*. Thus, *PR* is the most suitable pattern for user A in the independent drive situation, i.e., in the independent drive situation, user A prefers to participate in topics based on his personal interest factor. Similarly, as shown in Fig. 5(b), user B also has the $\max \ln L(S)$ in pattern *PR*. Therefore, both users exhibit the same behavior tendency in independent drive situation.



(a) (b) (c)

Fig. 7. (a) *EPR* for user B; (b) *IPR* for user B; (c) *E& PR* for user B.

Table 3
The driving strength of factors in six patterns of user B.

Situation	Model	Strength	maxlnL
Independent drive	ER	Se = 0.2650	−796.8900
	IR	Si = 0.3050	−807.9254
	PR	Sp = 0.6730	−603.9155
Combined drive	EPR	Se = 0.3140, Sp = 0.6620	−563.0384
	IPR	Si = 0.1040, Sp = 0.6720	−580.0551
	E&IPR	$S_{e\&i} = 0.2100$, Sp = 0.6680	−548.1718

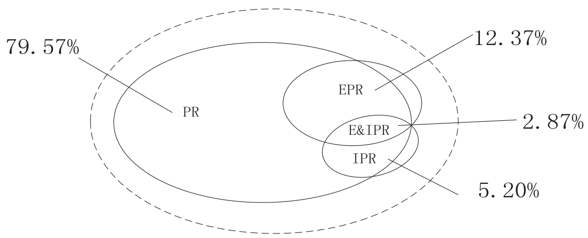


Fig. 8. The proportion of users preferring the different patterns.

The experimental result s for the two users in the combined drive situation are shown in Figs. 6 and 7. The abscissa and ordinate in the graphs represent the driving strength of two different driving factors. The contour indicates the value of the likelihood function.

As shown in Fig. 6(a–c), user A gets the higher $\max \ln L(S)$ in the *EPR* pattern. Thus, *EPR* is the most suitable pattern for user A in the combined drive situation, i.e., user A prefers to participate in topics based the explicit link factor and the personal interest factor. User B gets the higher maximized $\max \ln L(S)$ in pattern *E& IPR* as shown in Fig. 7(a–c). User B prefers to participate in hotspot topics based the explicit link factor, the implicit link factor and the personal interest factor.

Comparing the six patterns in the independent and in the combined drive situation, the most suitable pattern with the highest maximum value is capture. The driving strength in the different patterns and the corresponding values of the maximized log-likelihood are shown inTables 2 and 3.

For user A, *EPR* is the most suitable pattern. Consequently, user A much prefers to participate in topics based on the explicit link and his personal interest. The driving strengths of the explicit link factor and the personal interest factor are $S_e = 0.1180$ and $S_p = 0.7060$, respectively For user B, *E& IPR* is the most suitable pattern. User B prefers to participate in topics based on the explicit link factor, the implicit link factor and the personal interest factor. The respective driving strengths are $S_{e\&i} = 0.2100, S_p = 0.6680$.

Then, the users like A or B are classified according to the most suitable pattern, as shown in Fig. 8. First, the vast majority of users achieve higher maximized value s in the *ER* pattern. That means users tend to participate in hotspot topics that involve their personal interests. The *EPR* and *IPR* patterns are the most suitable patterns for 12.37% and 5.2% of of the users. Finally, 2.87% of the users are driven by explicit links, implicit links and personal interest together. In terms of a single driving factor, 8.07% of the users are driven by the implicit link factors, so the implicit link plays a very important role in driving users to participate in topics. Finally, we compare the driving strength of the same driving factors in patterns and *EPR*, *IPR* as shown in Fig. 9. The x-axis represents each user like user A or user B, and the y-axis represents the driving strength. The results indicate that personal interest as an internal factor has a greater driving strength, whereas external factors such as explicit and implicit links are relatively weak.

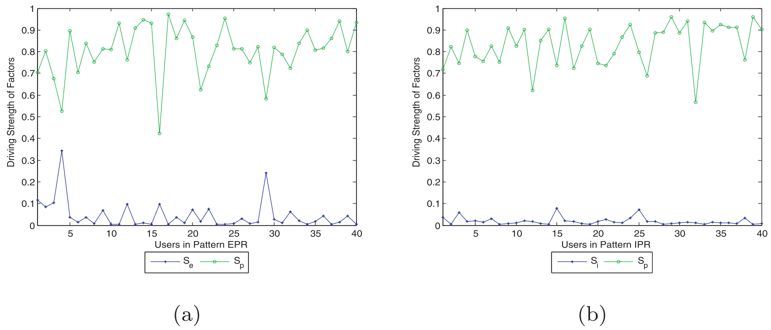


Fig. 9. a. The driving strength of factors in *EPR*; b. The driving strength of factors in *IPR* .

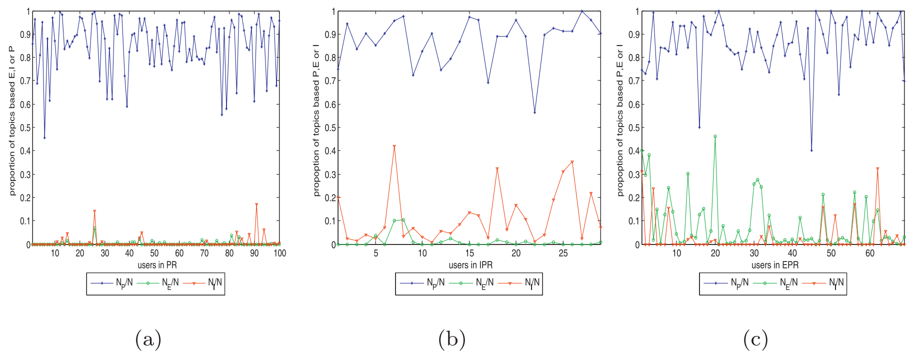


Fig. 10. a. Data statistics of users with *PR* pattern; b. Data statistics of users with *IPR* pattern; c. Data statistics of users with *EPR* pattern.

In order to verify the experimental results, we take the users in the *PR*, *EPR* and *IPR* patterns to carry out data statistics. The obtained statistical result (Fig. 10) allow us to verify whether the proportion of topics that users participate in based on the different factors is approximately the same as indicated by the results of the experiment. The x-axis in the graphs of Fig. 10 represents each user like user A or user B, and the y-axis represents the proportion of topics that user participated in based on the different driving factors. N_P is the number of topics the user participated in based on the personal interest factor. N_E is the number of topics the user participated in based on the explicit link factor. N_I is the number of topics the user participated in based on the implicit link factor. N presents the total number of topics user participated.

For the users with the *PR* pattern (Fig. 10(a)), the proportion of topics users participated based on personal interest is much higher than the proportion based on the implicit link factor and on the explicit link factor. For the users exhibiting the *IPR* pattern (Fig. 10(b)), the topics based personal interest still occupy a higher proportion. However, compared to Fig. 10(a), the proportions of topics based on the explicit link factor and on the implicit link factor are greater, and the proportion based on the implicit links factor is higher than that of the explicit links factor. Moreover, for the users with a *EPR* pattern (Fig. 10(c)), the proportion of topics based on the explicit link factor is obviously weaker than that of the implicit link factor.

The statistical analysis shows that, as an overall trend, the proportions of topics users participated in based on the different factors are consistent with their most suitable pattern. The experimental results of the statistical analysis.

6. Conclusion

In this paper, the user behavior is analyzed to mine the dynamic factors that drive users to participate in hotspot topics. In order to define the driving factor, we use a cloud model. Then, the *EIPR* model is established based on the theory of maximum likelihood estimation. Through the analysis of six patterns in this model, the factors that drive users to participate in hotspot topics are captured, and their driving strength are quantified. In addition, the role of implicit links in driving user behavior is investigated further.

According to the experimental result, different users are the influences by different factors in participating hotspot topics. Similarly, the driving strength s of the factors are also different. Furthermore, we divide the user group into different pattern groups according to the users' most suitable patterns. As shown in Fig. 8, implicit link factors drive 8.07% of the users to participate in topics. Therefore, implicit link factor plays a very important role in driving users to participate in hotspot topics. Analyz ing the driving strength of the same factor in each pattern groups. We determine that, as an internal driving factor, the personal interest factor has a higher driving strength, which is the main driver of user behavior. The driving strength of the external driving factors, such as the explicit link factor and implicit link factor, are weaker than the internal driving factor.

Acknowledgments

This work is supported by the National Key Basic Research Program (973 program) of China (No.2013CB329606), National Science Foundation of China (Grant No. 61272400), Chongqing Innovative Team Fund for College Development Project (Grant No.KJTD201310), Chongqing Youth Innovative Talent Project (Grant No. cstc2013kjrc-qnrc40004), Science and Technology Research Program of the Chongqing Municipal Education Committee (No.KJ1500425), WenFeng Foundation of CQUPT(No.WF201403) and Chongqing Graduate Research And Innovation Project (No.CYS14146).

References

[1] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The role of social networks in information diffusion, in: Proceedings of the 21st International Conference on World Wide Web, ACM, 2012, pp. 519–528.
[2] A.-L. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature 435 (7039) (2005) 207–211.
[3] M.S. Bernstein, E. Bakshy, M. Burke, B. Karrer, Quantifying the invisible audience in social networks, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2013, pp. 21–30.
[4] M. Cha, H. Haddadi, F. Benevenuto, P.K. Gummadi, Measuring user influence in twitter: the million follower fallacy., ICWSM 10 (10–17) (2010) 30.
[5] J.-F. Chen, J.-J. Yu, Y. Shen, Towards topic trend prediction on a topic evolution model with social connection, in: Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01, IEEE Computer Society, 2012, pp. 153–157.
[6] Y.-X. Dong, J. Zhang, J. Tang, N.V. Chawla, B. Wang, Coupledlp: link prediction in coupled networks, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 199–208.
[7] M. Gomez Rodriguez, J. Leskovec, B. Schölkopf, Structure and dynamics of information pathways in online media, in: Proceedings of the sixth ACM International Conference on Web Search and Data Mining, ACM, 2013, pp. 23–32.
[8] A. Guille, Information diffusion in online social networks, in: Proceedings of the 2013 SIGMOD/PODS Ph. D. symposium, ACM, 2013, pp. 31–36.
[9] W.-Y. Guo, S. Wu, L. Wang, T. Tan, Social-relational topic model for social networks, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM, 2015, pp. 1731–1734.
[10] Z. Jun-Fang, H. Xiao-Pu, W. Bing-Hong, Statistical property and model for the inter-event time of terrorism attacks, Chin. Phys. Lett. 27 (6) (2010) 068902.
[11] I. Karagiannis, A. Arampatzis, P.S. Efraimidis, G. Stamatelatos, Social network analysis of public lists of pois, in: Proceedings of the 19th Panhellenic Conference on Informatics, ACM, 2015, pp. 61–62.
[12] G. Katsimpras, D. Vogiatzis, G. Paliouras, Determining influential users with supervised random walks, in: Proceedings of the 24th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2015, pp. 787–792.
[13] K. Lee, J. Mahmud, J. Chen, M. Zhou, J. Nichols, Who will retweet this? detecting strangers from twitter to retweet information, ACM Trans. Intell. Syst. Technol. (TIST) 6 (3) (2015) 31.
[14] M.-J. Lee, C.-W. Chung, Finding k-highest betweenness centrality vertices in graphs, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2014, pp. 339–340.
[15] D. Li, Z.-M. Xu, S. Li, X. Sun, A. Gupta, K. Sycara, Link recommendation for promoting information diffusion in social networks, in: Proceedings of the 22nd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2013, pp. 185–186.
[16] D.-Y. Li, Artificial intelligence with uncertainty, in: Computer and Information Technology, 2004. CIT'04. The Fourth International Conference on, IEEE, 2004. 2–2
[17] H.-L. Li, C.-H. Guo, W.-R. Qiu, Similarity measurement between normal cloud models, Dianzi Xuebao(Acta Electronica Sinica) 39 (11) (2011) 2561–2567.
[18] L. Li, J. He, M. Wang, X. Wu, Trust agent-based behavior induction in social networks, Intell. Syst. IEEE 31 (1) (2016) 24–30.
[19] L. Liu, J. Tang, J.-W. Han, M. Jiang, S.-Q. Yang, Mining topic-level influence in heterogeneous networks, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, 2010, pp. 199–208.
[20] S. Ming-Sheng, C. Guan-Xiong, D. Shuang-Xing, W. Bing-Hong, Z. Tao, Interest-driven model for human dynamics, Chin. Phys. Lett. 27 (4) (2010) 048701.
[21] N. Ohsaka, T. Maehara, K.-i. Kawarabayashi, Efficient pagerank tracking in evolving networks, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 875–884.
[22] J.G. Oliveira, A.-L. Barabási, Human dynamics: darwin and einstein correspondence patterns, Nature 437 (7063) (2005). 1251–1251
[23] M.Z. Shafiq, M.U. Ilyas, A.X. Liu, H. Radha, Identifying leaders and followers in online social networks, Select. Areas Commun. IEEE J. 31 (9) (2013) 618–628.
[24] S. Shioda, Random walk based biased sampling for data collection on communication networks, ACM SIGMETRICS Perform. Eval. Rev. 42 (2) (2014) 21–23.
[25] J. Tang, S. Wu, J. Sun, Confluence: conformity influence in large social networks, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 347–355.
[26] G. Wang, H. Qing-Bo, P.S. Yu, Influence and similarity on heterogeneous networks, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, 2012a, pp. 1462–1466.
[27] H. Wang, A.A. Raza, Y. Lin, R. Rosenfeld, Behavior analysis of low-literate users of a viral speech-based telephone service, in: Proceedings of the 4th Annual Symposium on Computing for Development, ACM, 2013, pp. 1–9.
[28] M. Wang, W. Fu, S. Hao, D. Tao, X. Wu, Scalable semi-supervised learning by efficient anchor graph regularization, IEEE Trans. Knowl. Data Eng. (2016). 1–1
[29] S.-G. Wang, Q.-B. Sun, G.-W. Zhang, F.-C. Yang, Uncertain qos-aware skyline service selection based on cloud model, Ruanjian Xuebao/J. Softw. 23 (6) (2012b) 1397–1412.
[30] S.-Z. Wang, X. Hu, P.S. Yu, Z.-J. Li, Mmrate: inferring multi-aspect diffusion networks with multi-pattern cascades, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 1246–1255.
[31] Y. Wu, C. Zhou, M. Chen, J. Xiao, J. Kurths, Human comment dynamics in on-line social systems, Physica A 389 (24) (2010a) 5832–5837.
[32] Y. Wu, C. Zhou, J. Xiao, J. Kurths, H.J. Schellnhuber, Evidence for a bimodal distribution in human communication, Proc. Nation. Acad. Sci. 107 (44) (2010b) 18803–18808.
[33] Y.-P. Xiao, Research on Modeling and Algorithms of Human Dynamic in Online Social Networks, Beijing University Of Posts And Telecommunications, 2013 Ph.D. thesis.
[34] Q. Yan, L. Yi, L. Wu, Human dynamic model co-driven by interest and social identity in the microblog community, Physica A 391 (4) (2012) 1540–1545.
[35] C.-H. Yang, D.-Y. Li, Planar model and its application in prediction, Chin. J. Comput. 21 (11) (1998) 961–969.
[36] R. Yang, L. Zhuhadar, Extensions of closeness centrality? in: Proceedings of the 49th Annual Southeast Regional Conference, ACM, 2011, pp. 304–305.
[37] Y. Yang, Y.-X. Dong, N.V. Chawla, Predicting node degree centrality with the node prominence profile, Sci. Rep. 4 (2014).
[38] G.-W. Zhang, D.-Y. Li, P. Li, J.-C. Kang, G.-S. Chen, A collaborative filtering recommendation algorithm based on cloud model, Ruan Jian Xue Bao(J. Softw.) 18 (10) (2007) 2403–2411.
[39] T. Zhou, H.A.-T. Kiet, B.J. Kim, B.-H. Wang, P. Holme, Role of activity in human dynamics, EPL (Europhys. Lett.) 82 (2) (2008) 28002.