



Event Detection and Multi-source Propagation for Online Social Network Management

Lei-lei Shi¹ · Lu Liu² · Yan Wu¹ · Liang Jiang¹ · Ayodeji Ayorinde³

Received: 10 September 2018 / Revised: 31 January 2019 / Accepted: 18 February 2019 /
Published online: 13 March 2019
© The Author(s) 2019

Abstract

The social network is a huge source of information, which plays an increasingly crucial role in people's daily lives. As a form of online social network management, much information can be discovered via posts, which allows people to exchange and propagate real-life events. Multi-source event propagation involves relevant posts of interesting topics from some key users to others in microblogging network users for network management. However, there are many noisy data in traditional microblogging network management. Meanwhile few people study the spontaneous transmission of events in microblogging network management, as well as the cooperation and competition among multiple event sources. To this end, the event detection and multi-source propagation model, is established. Specifically, for efficient and accurate result of the hot event detection and propagation, we obtain the information of previous event detection and propagation to create some experience sets for the intelligent event propagation. And a multi-source events propagation model based on individual interest is established to describe the process of multi-source event information detection and dissemination, and to describe the key role of users and information characteristics in the process of communication and network management. Meanwhile, the experimental results show that the proposed intelligent multi-source events detection and propagation model can learn from previous propagation to better discover and propagate the hot events under users' changing interest. Besides, the interaction broadens the influence scope of hot events. This helps to explain the formation of microblogging hot events dissemination, to provide a theoretical basis for the research and network management of the guiding strategy.

Keywords Microblogging · Intelligent · Multi-source · Events propagation · Network management

✉ Lu Liu
ll297@leicester.ac.uk

Extended author information available on the last page of the article

1 Introduction

In recent years, online social network management has become an important part of our daily lives [1–5]. As a form of online social network management, microblogging network management platforms are also developing and attracting people at a rapid pace [6–10]. Microblogging network management platforms is known as the best tool for people to share and exchange opinion [11–14]. For example, many companies can promote their goods and services via microblogging network management platforms. Some people who have interest in football games can get information of their favorite football players immediately via relevant posts of users on microblogging network management platforms, which serves as tools for sending posts, which will also allow users to discover trending events [15–19].

On microblogging network management platforms [20, 21], the spread of information is likened to fission. After the release of the user's post, the microblogging network management platforms will automatically push these posts to neighbours. These neighbours may forward these posts, which will be pushed to the neighbour's neighbours. And, the user is not only the consumer of information, but also the producer of information. Users forward other people's posts, but also release new information. The information can also be forwarded by neighbours, thus spreading to more users. Therefore, in a microblogging environment, the information spread faster and local discussion is more likely to cause group effect. At the same time, users of microblogging network management platforms obviously have different participation behaviours, different interests in topics, different levels of activity, and the contents of post affects their behaviour, which results in heterogeneity of topics. In addition, most of the topics will quickly disappear from the list of discussed topics, and some of the topics will stand out amongst competing topics to become a hot topic, causing a lot of attention.

At present, the dynamics of information dissemination is like infectious disease dynamics, such as the SIR model [22]. These models assume that there is only one communicator in the system at the initial time, and the communicator will pass the information to the neighbours through interaction with the neighbours. At the same time, interest in the communicated information may slow down leading to a loss enthusiasm from users and exit the topic discussion pool, to enter the stable state. However, users in microblogging network may spontaneously publish new posts and become a communicator. At the same time, users may not be interested in these events information, and will not get involved in the communication. The existing information propagation models [22–25] do not consider multi-source events detection and propagation, competing hot events and user interaction, hence it is needed to establish a multi-source events detection and propagation model which is suitable to describe the process of hot events information dissemination, and to describe the key role of users and event characteristics in the process of communication.

To this end, the multi-source events detection and propagation model, named event detection and multi-source propagation (EDMP) model, is proposed. And

we study the propagation process of a single hot event, modelling the individual spontaneous communication behaviours. Then, interaction between the communication sources is analysed in the model. Finally, we study the process of simultaneous communication and establish multi-source propagation model for competing events based on user interest, which describes the relationship between the hot events.

The main contributions of this paper are listed as follows:

1. We propose an intelligent event propagation model with the knowledge sets [8]. Specifically, this event propagation model does not require any knowledge from the microblogging network. The microblogging network management platform only assigns a set of key users as initial users representing hot events to the event propagation model. Then, the model will use the first initial users set for learning and generating experience sets. And the event propagation model will compare the keywords of users' interest with the content of discovered hot events, and it should have proper keywords to describe the topic of users' interest for comparison to form the topic keywords experience set. Therein, the model will compute a prediction score according to the information of users and events obtained from previous event propagation. What's more, the prediction score will be used in user ordering process to generate target users learning experience set when the next event propagation starts. Finally, for the next hot event propagation, it will use these experience sets to achieve more effective event propagation.
2. We establish multi-source events detection and propagation model based on individual interest [7] to describe the process of multi-source event information dissemination, and to describe the key role of users and hot event characteristics in the process of microblogging network management and communication. Specifically, we firstly study the propagation process of a single event, modelling the individual spontaneous communication behaviours. Each time, an individual chooses a message to participate in from the information disseminated from the information collection, at the same time consider their own topics of interest. Then as in the model, there is a cooperative relationship among the communication sources, we finally study the process of simultaneous communication and establish a multi-source event propagation model, which describes the relationship between the hot events based on cooperation and competition.
3. We apply our model to the real Twitter dataset to demonstrate the effectiveness of our proposed multi-source events detection and propagation model compared with some existing event detection and propagation models [7, 14, 17, 26].

The remainder of this paper is structured as follows: we discuss related work for event detection and propagation in Twitter in Sect. 2. In Sect. 3, we introduce our intelligent event propagation model. In Sect. 4, we design our multi-source events propagation model. We do our experiments in Sect. 5. The last section concludes our study and future work.

2 Related Work

Recently, event detection and propagation, has drawn more and more attention from various fields of research especially concerning influence maximization based on users' opinion, and all kinds of methods have been proposed to catch event propagation in social networks [13, 14, 22–25]. Besides event detection and propagation have extensive applications such as viral marketing [1], product promotion [2], and friend recommendation [11] and rumors control [12]. And some researchers pay attention to create some effective models for explains the general process of event information dissemination. These models are useful for the dissemination of event information in social network simulation [24, 25, 27]. However, these models cannot be directly applied to propagation of hot events because of the complex processes involved and uncertainty.

As we all know, the influence maximization problem was first introduced into the social network as an algorithm proposed by Richardson, which has been proved to be NP-hard, and can perform an approximate optimal solution with the accuracy of $(1 - 1/e)$ based on a greedy algorithm. Developments from this initial work have generated excellent algorithms [28–30] and effectively improve the time efficiency of mining influential nodes. However, the influence of a given node on other nodes is the same in those studies; that is to say, the activation chance of a node to activate other nodes is a constant. Similarly, information content and node preference are not taken into consideration; i.e. the influence exerted by a node is also fixed, even for totally different event topics. Obviously, this is not accurate in real life, where, for example, an individual may has high influence among peers when it comes to discuss the subject “economy” but completely unknown by peers in the area of “law”. In simple terms, it is rare for any individual to be considered as an expert in multiple fields. The influence of a person in a social network is likewise related to both the node and the topic, and the influence of a given node is different for different topics [31–33]. However, a limitation of these works [31–33] is that it solely considered the topic influence on user activation probability and did not take into account the popularity degree of the users' interest, the links between posts and the diffusion power of users. Overall, this results in a low efficiency algorithm and the improper number of final mining core users. Minimal research combines the topic popularity degree scoring, topic community detection and event propagation together, which can improve the efficiency and enlarge the influence scope of key users of hot events.

Therefore, previous researches have focused on studying event propagation in various ways [34–39]. Richardson and Domingos [32] studied the information propagation problem and propose a probabilistic method. And Kempe et al. [27] formulated the problem of event propagation as an optimization problem and developed an algorithm for an event diffusion model. Meanwhile, some other researchers have also put forward a lot of excellent algorithms on the basis of this work [22, 24, 25] and effectively improve the time efficiency of event propagation. However, the event propagation model on all other users is the same in those studies, that is, the activation probability of a user to activate other users is a constant. Each of these studies does not consider the importance of event content and user preferences. It has been

shown that the event propagation in the microblogging network is related to the relationship between the users and the topic of event, and the event propagation ability of the same user is different under different events [27, 40, 41].

To this end, Zhang et al. [42] proposed a two-stage algorithm to propagate the hot events for a specific topic and improves the event propagation scope. Zhou et al. [43] calculated the user activation probability at the topic-level by user interest distribution and then proposed a new event propagation algorithm to quickly diffuse the events under specific topic based on the probability, which also improves the event propagation scope. These studies focus on the effect of event influence on user's activation probability and do not consider the popularity of events, the links between posts and the diffusion power of users, which also causes the waste of users' influence, resulting to low efficiency of event propagation. Meanwhile, unlike our model, these works only studied single event propagation. In addition, few people study the spontaneous transmission of events in microblogging network, as well as the interaction and competition among event sources. To the best of our knowledge, the intelligent multi-source events propagation question has not yet been well discussed.

Although previous researches have proposed many methods to event propagation, our work is very different. First, we propose the new event detection and propagation model based on key users [19] and users' interest [7]. Meanwhile, we express the problem of event propagation as a learned task and aim to identify the accurate characteristics of such events. Then we investigate the relation between extracted features of event propagation and user interest. Last, our dataset is extracted from Twitter, and we validate the effectiveness of our model compared with existing models [7, 14, 17, 26].

3 Intelligent Events Propagation Process

3.1 Preliminary

Given a microblogging network $G=(V,E)$, $V=\{v_1,v_2,\dots,v_n\}$ is a set of users, $E=\{e_1,e_2,\dots,e_m\}$ is a set of edges. Adjacency matrix denotes the connection relationship among users, the value of the corresponding element of the matrix indicates whether the edge exists: if there is an edge between v_i and v_j , then $A_{ij}=1$; if no edge exists between v_i and v_j , then $A_{ij}=0$.

Generally, the adjacency matrix A can be used as the similarity matrix of the microblogging network to describe the similarity between users. However, in addition to the similarity between the users which are directly connected in the network, there are different degrees of similarity between the users which are not directly connected. For example, there is a certain similarity between two users that can reach one another after a finite number of steps. The adjacency matrix is used as the similarity matrix of the network and it can simply represent the similarity relationship between users that are directly connected but it cannot be used to express the similarity relationship between the users that are not directly connected. Therefore, the adjacency matrix loses the similarity relation information between many users

and cannot reflect the complete local information of each user. Adjacency matrix contains limited information which affects the accuracy of community discovery.

Therefore, in order to describe the local information of each user more adequately, a method based on step number is proposed in this section. According to the adjacency matrix A in the network, the similarity relationship score between users is calculated, and a new similarity matrix is obtained. The definitions of s -steps and similarity matrix are given in this paper as follow.

Definition 1 (*s-steps*) Given a social network $G=(V,E)$, For any user in the point set, if user u can arrive at user v at least after s steps, that is, the length of the shortest path from user u to user v is s , it will be said that user u can arrive at user v through s steps.

Step number and attenuation factor are used to calculate the similarity relationship between two users which are not directly connected, which can better reflect the community topology structure, and improve the accuracy of community detection [15]. However, when the number of steps is greater than a certain threshold, two users that are not in the same community will also get a certain similarity value, which makes the boundary of community structure more obscure. Therefore, setting step threshold S , only calculates the similarity between users that can reach each other in the S steps, so as to ensure that the topological information of the micro-blogging social network is enhanced without affecting the division of community boundaries. In the experimental part, the step number threshold S and attenuation factor σ are analyzed, and the influence of different step threshold S and attenuation factor σ on the result is studied.

3.2 The Improved HITS Method

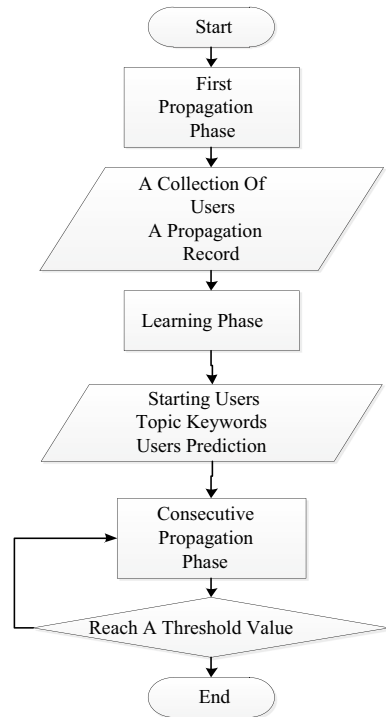
In the original HITS method, a link is used to represent the hyperlinks between web pages. While in our improved HITS method, a link represents an operational relationship between a user and a post such as publishing or commenting.

In this paper, the HITS algorithm is extended to exploit the inseparable connection between the users and their corresponding posts for the purpose of distilling the influential users [7, 17, 19]. As a result, the proposed improved HITS method can effectively filter out the random ordinary users, this helps to improve the efficiency and accuracy of intelligent event propagation model.

3.3 Intelligent Event Propagation Process

As we can see from the Fig. 1, it depicts the process of the intelligent event propagation. Intelligent event propagation consists of three steps: first propagation, learning process and consecutive propagation, in which the learning process is pretty important because the intelligent event propagation model's experience sets will be gained from this process.

Fig. 1 The diagram of intelligent event propagation process



First propagation is a step of propagating events without any prior information about how to choose the initial users. During this step, the event propagation model only has some keywords extracted from key posts describing an interesting topic from an event. The key users are chosen to be the candidate set of initial influential users for propagating the hot events.

Learning process is a step where the event propagation model learns how to better get the relevant influential users. First, initial influential users set will be obtained by computing a hub score for each user and obtaining the high hub ones based on the HITS algorithm. Besides, topic keyword set will be created by extracting keywords from users' interests [7], as well as from the key posts of users, which point to hot events [17]. Finally, target user prediction set will be achieved by calculating topic similarity between the content of all detected hot events and the content of all detected users' interests [7, 9, 10] and employing those scores in user prediction process. These sets are composed of the intelligent event propagation model's experience sets. As we all know, appropriate initial influential users support the model to propagate as many influential users of hot events as possible at the beginning process of hot event propagation. What's more, proper topic keywords will help the model to recognize from the propagated users, the keywords related to a topic of users' interest. Furthermore suitable target user prediction assists the model to predict the relevancy of the content of users extracted from hot events.

Consecutive propagation is a step during which the event propagation model detects high influence users based on these experience sets. During this process, suitable initial influential users and high-quality topic keywords have been learned.

3.4 Topic Popularity Based Event Propagation

In the IC model, the activation probability is generated randomly. However, the activation probability of a node is related to the social relationships among nodes and topics in the process of event propagation, and nodes have a different activation probability for different topics. Therefore, a Topic Popularity-based Event Propagation model, named TPEP model is proposed, which calculates the node activation probability $P_{u,v}^t$ for specific topics to simulate the event propagation in social networks in a more realistic way.

The activation probability $P_{u,v}^t$ is influenced by the following factors. Firstly, it is closely related to the social connections between nodes; greater connection times imply a more intimate relationship between nodes and have a higher activation probability. Therefore, user intimacy can be used to represent the degree of intimacy between nodes.

Definition 2 User intimacy, $C_{u,v}$, denotes the frequency of the connection between nodes u and v . It can be obtained from the ratio of the connection times of u and v to the connection times of u and other nodes. The calculation method is shown in formula (1).

$$C_{u,v} = \frac{R_{u,v}}{\sum_{i=1}^n R_{u,V_i} + \sum_{i=1}^n R_{v,V_i}}, (u, v, V_i \in V) \quad (1)$$

where R_{u,V_i} denotes the connection time of nodes u and V_i , $R_{u,v}$ denotes the connection time of nodes u and v .

In addition, $P_{u,v}^t$ is also influenced by users' topic popularity. The more popular the two users' topics are, the easier and quicker information is propagated. Therefore, the topic popularity can affect the activation probability $P_{u,v}^t$ of two users.

Definition 3 Topic popularity, $TP_{u,v}^T$, denotes the popularity degree of two users' topic. The topic popularity $TP_{u,v}^T$ can be calculated as formula (2).

$$TP_{u,v}^T = \frac{Authority_{u,v}^T}{Authority_{\max}^{T_i} + Authority_{\min}^{T_i}}, (u, v \in V, T_i \in T) \quad (2)$$

where $Authority_{u,v}^T$ denotes the authority of key post in the topic T , $Authority_{\max}^{T_i}$ denotes the biggest authority of key posts in topics and $Authority_{\min}^{T_i}$ denotes the smallest authority of key posts in topics. The authority of key posts in topics can be gained computed on improved HITS algorithm which is depicted above. Besides

if the authoritative value of key posts occupies a significant part of the topics, the more popular the topic will be.

In summary, the activation probability $P_{u,v}^t$ is influenced by the user intimacy $C_{u,v}$ and topic popularity $TP_{u,v}^T$, so the activation probability of user u to v for specific topic t is calculated using formula (3).

$$P_{u,v}^t = C_{u,v} \times TP_{u,v}^T \quad (P_{u,v}^t \in [0, 1]) \quad (3)$$

The propagation process of the TPEP model is the same as the IC model that each user has only one chance to activate its neighboring users, and the user's activation process is independent of each other. The difference is that the users' activation probability of TPEP is different under different topics, which is more in line with the information propagation of microblogging networks.

In the first stage, we only choose the initial influential spreaders, and not consider the information propagation characteristics of the microblogging network. Therefore, this second stage uses the spreaders from the first stage to spread information using the TPEP model proposed in this paper, it then iteratively mines top-k spreaders with biggest topic influence increment as the remaining influential nodes. The biggest topic influence increment refers to the influence scope value of the spreader set after adding a spreader u minus the scope value before adding the spreader u to achieve a maximum. The calculation method is shown in formula (4).

$$\delta(u|t) = \max\{\delta(S \cup \{u\}|t) - \delta(S|t)\} \quad (4)$$

4 Experiments

In this section, we detail the experiments in order to show the effectiveness of our proposed EDMP model. We consider typical event detection and propagation models as our baseline, namely IC (Independent Cascade) [14], BEE (Bursty Event dEtection) [26], EVE (Efficient eVent dEtection) [17], HEE (Hot Event Evolution) [7].

4.1 Dataset

Our datasets are collected from Twitter (<http://twitter.com/>) via Twitter API [20]. The collected dataset is composed of 1,500,000 posts and 36,845 users.

4.2 Baseline Approaches

The efficiency and effectiveness of the proposed EDMP model is validated by evaluating our model against IC model, BEE, EVE, HEE, which are the classic event detection and propagation algorithms.

4.3 Parameter Experiment

The effect of step number threshold S and attenuation factor σ on experimental results are in this section. 1000 users' data in the database are randomly selected for experiments, and the F -measure score mentioned above is a measure of the index.

In the experiment, the value of one parameter is fixed, and the influence of the change of the other parameter value on the F -measure is analyzed to determine the final value of the parameters.

(1) Step number threshold S

In view of the data set, the attenuation factor $\sigma=0.5$ is set up, and the effect of the step number threshold S on the F -measure is analyzed.

As shown in Fig. 2, with the increase of the step number threshold S , the trend of F -measure increases first and then decreases. The experimental results show that considering the similarity of user pairs which are not directly connected but reachable within a certain number of steps, the local information structure of each user can be effectively determined. However, if the threshold is too large, the distance between the users in the same community will also increase with a certain similarity value, which will not facilitate the identification of the community boundaries, and the accuracy of the community will be reduced. For small datasets, select small step number threshold 3, and for big datasets, select slightly larger step threshold 8 to achieve the optimal result. The threshold selection in this paper is 3.

(2) Attenuation factor σ

In view of the data set, step number threshold $S=0.5$ is set up, and the effect of the attenuation factor σ on the F -measure is analyzed.

As shown in Fig. 3, with the increase of attenuation factor, the trend of F -measure overall increases first and then decreases. This due to the fact that the attenuation factor controls the attenuation degree of similarity with the increase of hop counts. For small datasets, a slight attenuation factor $\sigma=0.5$ is selected to avoid the

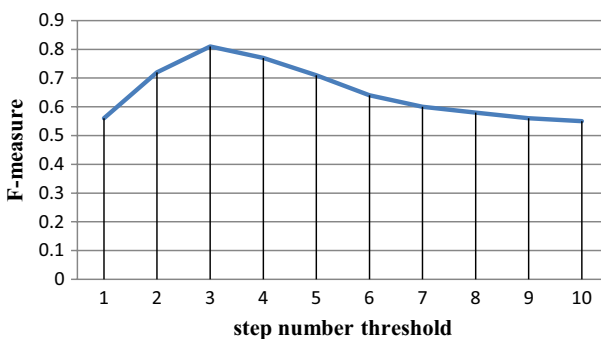


Fig. 2 F-measure score under the different step number threshold

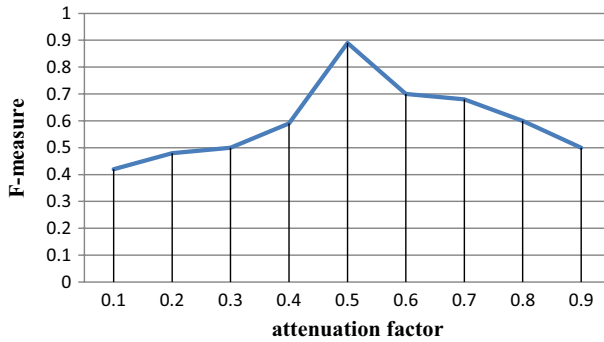


Fig. 3 F-measure value under the different attenuation factor

vagueness of community boundary when the attenuation factor is too large. For a large dataset, a small attenuation factor $\sigma=0.1$ is selected to enhance the local feature of the user to achieve the optimal result.

4.4 Evaluation

The Precision is an important metric, which can be used to measure the efficiency of our proposed model, as defined as follows:

$$\text{Precision}_p = \frac{k}{K} \tag{5}$$

where k represents the number of posts related to the real-life event in the top K posts under a topic.

As mentioned above, the scoring method based on HITS algorithm is proposed to select high-quality posts, high-influence users and high-popularity topics from the social media data streams. Threshold A is then defined and posts (where the authority score are greater than A) are high-quality.

Three experiments are conducted setting different value to get a suitable threshold A . Table 1 shows the result of the number of detected hot events under different topic count m respectively. Tables 2 and 3 show the result of time efficiency and precision. Three experiments testified that EDMP model can detect

Table 1 Number topic of different value

Value	Number of detected hot event under the different m		
	$m=10$	$m=15$	$m=20$
$A=0.0001$	7	11	16
$A=0.001$	6	8	11
$A=0.01$	5	7	9

Table 2 Time of different value

Value	Time		
	Event detection (min)	Event propagation (min)	Total (min)
A=0.0001	20.6	17.8	38.4
A=0.001	20.6	22.5	43.1
A=0.01	20.6	17	37.6

Table 3 Precision of different value

Value	P@10	P@20	P@50	P@100
A=0.0001	10/10	19/20	45/50	56/100
A=0.001	10/10	17/20	41/50	53/100
A=0.01	10/10	17/20	43/50	46/100

hot events more accurately and efficiently when $A=0.0001$. Therefore, the next contrast experiments are all conducted with $A=0.0001$.

We present a propagation result on a two-dimensional graph in Fig. 7 where x -axis is the number of propagated users and y -axis is a precision obtained as follows.

$$\text{Precision}_u = \frac{\text{number of relevant users at that time}}{\text{number of total relevant users}} \quad (6)$$

In our experiments, we set the top 10 popular events to be our multi-source events set for showing the performance of our proposed EDMP model. At the same time, we will focus on the top 10 users for each propagation process and calculate their influence scope.

1. Filtering the hot events based on topic decision model: We can also detect the proper number of hot events from Fig. 4 according to the number of key posts, which also plays a key role in the spread of influence under a specific user interest community. And it can be seen from Tables 4 and 5, our proposed EDMP model can detect the top k (k is set to 10 in Table 4) high-quality posts according to their authority value efficiently and effectively. When the authority value of posts is equal, it can be sorted according to the minimum distance of the key posts.

2. The initial starting users for the first propagation: As is shown in Table 6, we can see the degree and hub value of users for topics, which can distinguish the importance of users under each popular topic. Meanwhile, we can also discover the number of influential users for each popular topic from Table 6, by setting different number of initial influential users. With the increase of the number of initial influential users, the influence scope is achieved to 82 when the number of initial influential users is 10 and remains the same later from Fig. 5. And the top 10 initial influential spreaders and the popular topics they belong to are shown from Table 7, which plays a key role to the spread of influence for specific users' interests.

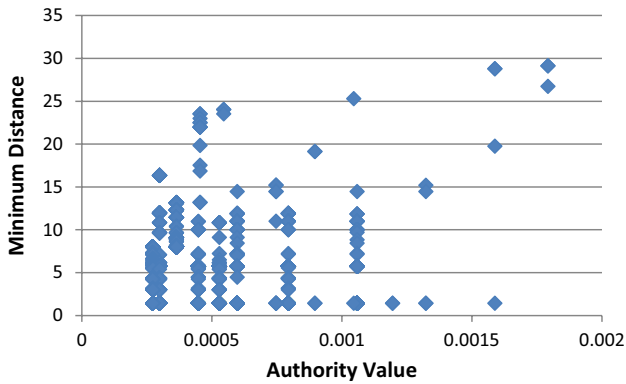


Fig. 4 The number of hot events

Table 4 Minimum distance and authority of posts

Post ID	Authority value	Minimum distance
681693469564383232	0.001792382	29.12043956
681697568456192001	0.001792382	29.12043956
681699684168015873	0.001588142	28.7923601
681697799730249728	0.001588142	28.7923601
681697033523077122	0.001045556	26.73948391
681695337304702976	0.000896191	25.29822128
681697928268910593	0.000545355	24.0208243
681696803629219840	0.000545355	24.0208243
684205783525888002	0.000545355	23.53720459
681695402928648193	0.000454463	23.53720459

Table 5 Key posts under popular interests

Post ID	Popular interest
681693469564383232	Sport
681697568456192001	Sport
681699684168015873	Sport
681697799730249728	Sport
681697033523077122	Sport
681695337304702976	Music
681697928268910593	Music
681696803629219840	Music
684205783525888002	Economy
681695402928648193	Emotion

Table 6 Degree and hub value of top 10 influential users under topics

User ID	Hub value	Degree	Interest
339283603	0.003429355	24535	Sport
1679619506	0.003233392	2869	Sport
3693887599	0.003135411	334	Music
933364430	0.002253576	1157	Sport
4068440360	0.00186165	377	Emotion
1000421510	0.001665687	1458	Music
2168821905	0.001567705	21973	Emotion
3254047099	0.001567705	489	Emotion
2310175028	0.001273761	1778	Music
863205451	0.000979816	44	Conflict

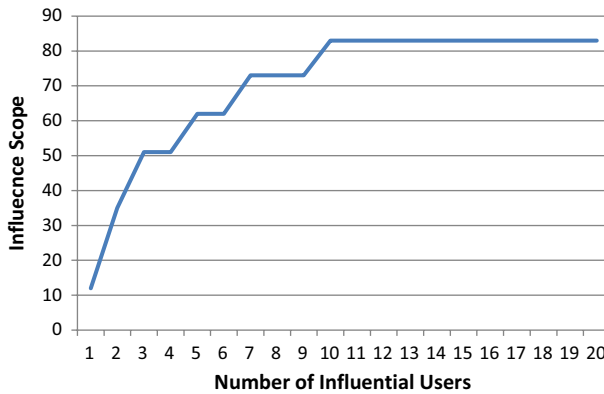


Fig. 5 The influence scope of IC model

Table 7 Top 10 initial influential spreaders mining and the popular topics they belong to

User ID	IF	Popular interest
339283603	0.051440325	Sport
1679619506	0.03233392	Sport
3693887599	0.03135411	Music
933364430	0.020282184	Sport
1000421510	0.01303155	Music
4068440360	0.011757792	Emotion
1367531	0.011757792	Economy
3254047099	0.011659809	Emotion
2310175028	0.010973935	Music
2168821905	0.010973935	Emotion

Table 8 The final influence scope of event detection and propagation

Event propagation	BEE + IC	EVE + IC	HEE + IC	EDMP
The first propagation	78	78	82	82
The second propagation	92	95	96	110
The third propagation	82	81	84	226
Consecutive propagation				230

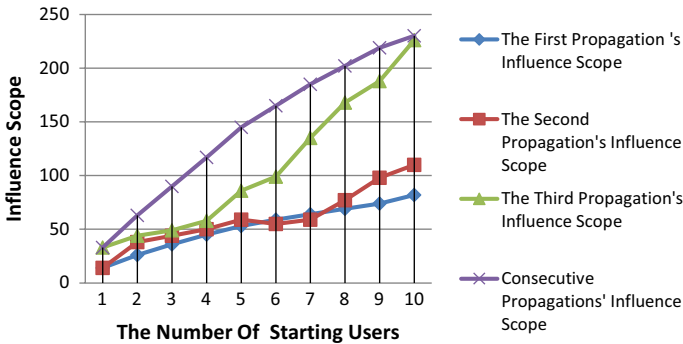


Fig. 6 The influence scope of propagation for EDMP model

3. The contrast of final influence scope results about initial users' discovery: In order to verify the effectiveness of influence scope of the proposed EDMP model, all four algorithms are running on the same configuration of PC. The experiment was repeated 5 times to compute the average value, then comparing the influence scope of users discovered by these four models, the experimental results are shown in Table 8 and Fig. 6. We can see that the proposed EDMP model outperforms the other three IC based models. This is because the proposed EDMP model considers the impact of the topic popularity, and it selects enough number users with high topic diffusion power as the influence users where their influence scope spreads most of the topic areas. Besides, the proposed EDMP model builds three kinds of knowledge sets, i.e. starting users, topic keywords and target users' prediction. These knowledge sets are outputs of the intelligent event propagation model's learning process. Proper initial users support the model ability to select as many influential users as possible at the beginning of event propagation process. Suitable topic keywords help the model to recognize, from the gathered users, the keywords related to a topic with considerable users' interest. Good target user prediction assists the model to predict the relevancy of the content of user's key posts extracted from the hot events. However, the BEE + IC model and EVE + IC model do not considered the topic diffusion power of the users and the popularity of topics, so the number of selected users is not adequate under specific event. Meanwhile, HEE + IC model do not take into account the learning ability of consecutive propagation, thus the influential spreaders discovered by this paper are the most adequate set compared with BEE + IC, EVE + IC and

HEE + IC models. This is because the activation probability of the IC model is not stable and the propagation of IC model is one event. However, our presented EDMP model can improve its initial users through three knowledge sets.

4. **Learnable Ability and Precision Analysis of Multi-source Events propagation:** We first set initial topics set of events as ‘Basketball’, ‘Music’, ‘Economy’ and ‘Emotion’ to describe the multi-source events. We then start the event propagation for selecting top 10 users to be the proper number of initial set of starting users. The first event propagation process will be used to build the three experience sets. Besides, each consecutive event propagation process has been done using experience sets built and learned from the previous event propagation process. Finally, Fig. 7 shows the learnable capability of the EDMP model for the first, the second and the third propagation process.

When we investigated the interests of users found in the topic keywords experience set, we found that the EDMP model can incrementally learn new interests of users from the previous event propagation process which can extract the set of users’ topic of interest, such as ‘Basketball’, ‘Music’, ‘Economy’ and ‘Emotion’, i.e. it could use ‘Basketball’, ‘Music’, ‘Economy’ as a set of users’ topic of interest in the second propagation process and use ‘Basketball’, ‘Music’ as a set of users’ topic of interest in the third propagation process. This is because the proposed EDMP model builds three kinds of experience set, i.e. starting users, topic keywords and target users prediction. These experience set compose the intelligent event propagation of the model’s learning experience. Proper starting users help the model to identify as many relevant users as possible at the beginning of propagation process. Appropriate topic keywords help the model to recognize, from the gathered users, the keywords related to a topic of users’ interest. Suitable target user prediction assists the model to predict the relevancy of the content of users extracted from a hot event.

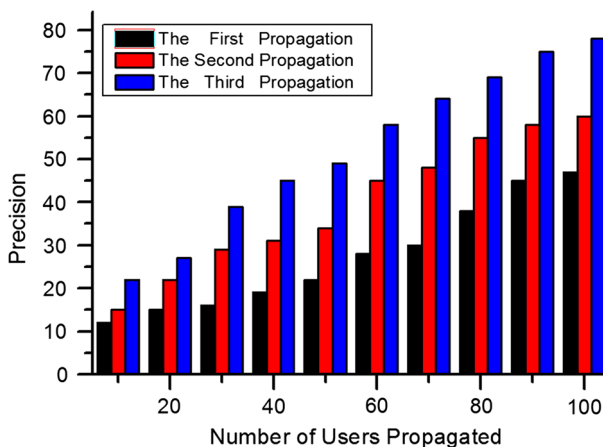


Fig. 7 The learnable ability results in improvement of the precision of users propagated during the consecutive propagation process

5 Conclusion and Future Work

In this paper, we present a novel approach to build an intelligent event propagation model which is capable of learning from event propagation experience and adapts itself to better propagation through relevant users and key posts during consecutive propagation process for microblogging network management. Specifically, for efficient and accurate result of the next event propagation, we derive the information of previous event propagation process to build some experience sets: starting users, topic keywords and target users' prediction. These experience sets are used to build the experience sets of the intelligent event propagation model to produce better result for the next propagation. And we study the propagation process of a single hot event, modelling the individual spontaneous communication behaviours. Then, an interactive relationship among the communication sources is analysed in the model. Finally, we study the process of simultaneous communication, and establish multi-source events competition propagation model based on user interest, which describes the relationship between the hot events.

Besides, the competitions between events shorten the survival time, and at the same time, the cooperation broadens the influence scope of hot events. This help to explain the formation of microblogging's hot events dissemination, to provide a theoretical basis for the research of the guiding strategy about the online social network management. Meanwhile, the next research points will be how to predict the links of target users during the event propagation and how to predict the users' behaviour evolution in hot events propagation process in the future.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China under Grants No. 61502209 and 61502207, Natural Science Foundation of Jiangsu Province under Grant BK20170069, and UK-Jiangsu 20-20 World Class University Initiative programme.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Gao, Q., Abel, F., Houben, G.J., Yong, Y.: A comparative study of users' microblogging behavior on sina weibo and twitter. In: International Conference on User Modeling (2012)
2. Diao, Q., Jiang, J., Zhu, F., Lim, E.-P.: Finding bursty topics from microblogs. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, vol. 1, pp. 536–544 (2012)
3. Wu, Y., Yan, C., Liu, L., Ding, Z.J., Jiang, C.J.: An adaptive multilevel indexing method for disaster service discovery. *IEEE Trans. Comput.* **64**(9), 2447–2459 (2015)
4. Liu, L., Antonopoulos, N., Xu, J., Webster, D., Wu, K.G.: Distributed service integration for disaster monitoring sensor systems. *IET Commun.* **5**(12), 1777–1784 (2011)
5. Panneerselvam, J., Liu, L., Antonopoulos, N., Yuan, B.: Workload analysis for the scope of user demand prediction model evaluations in cloud environments. In: Proceeding of 7th IEEE International Conference on Utility and Cloud Computing (UCC 2014), London, pp. 883–889 (2014)

6. Bao, J., Zheng, Y.: Location-based and preference-aware recommendation using sparse geo-social networking data. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, pp. 199–208 (2012)
7. Shi, L.L., Liu, L., Wu, Y., Jiang, L., Hardy, J.: Event detection and user interest discovering in social media data streams. *IEEE Access* **5**, 20953–20964 (2017)
8. Aldhaferi, A., Lee, J.: Event detection on large social media using temporal analysis. In: Computing and Communication Workshop and Conference (2017)
9. Jiang, L., Shi, L.L., Liu, L., Yuan, B., Zheng, Y.J.: An efficient evolutionary user interest community discovery model in dynamic social networks for internet of people. *IEEE Internet Things J.* (2019). <https://doi.org/10.1109/JIOT.2019.2893625>
10. Jiang, L., Shi, L.L., Liu, L., Yousuf, M.A., Yao, J.J.: User interest community detection on social media using collaborative filtering. *Wirel. Netw.* (2019). <https://doi.org/10.1007/s11276-018-01913-4>
11. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, vol. 8, pp. 50–57 (1999)
12. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
13. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.: The role of social networks in information diffusion. In: Proceedings of the 21st International Conference on World Wide Web, ACM, pp. 519–528 (2012)
14. Guille, A., Hacid, H., Favre, C., Zighed, D.A.: Information diffusion in online social networks: a survey. *ACM Sigmod Rec.* **42**, 17–28 (2013)
15. Li, Y., Jia, C., Yu, J.: A parameter-free community detection method based on centrality and dispersion of nodes in complex networks. *Phys. A Stat. Mech. Appl.* **438**, 321–334 (2015)
16. Yan, P.: Mapreduce and semantics enabled event detection using social media. *J. Artif. Intell. Soft Comput. Res.* **7**, 201–212 (2017)
17. Sun, X., Wu, Y., Liu, L., Panneerselvam, J.: Efficient event detection in social media data streams. In: IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 1711–1717 (2015)
18. Zhou, Y., Xu, H., Lei, L.: Event detection based on interactive communication streams in social network. In: EAI International Conference on Mobile Multimedia Communications, pp. 54–57 (2016)
19. Shi, L., Wu, Y., Liu, L., Sun, X., Jiang, L.: Event detection and identification of influential spreaders in social media data streams. *Big Data Min. Anal.* **1**, 34–46 (2018)
20. Twitter, REST API v1.1 Resources, 2019. <https://dev.twitter.com>. Accessed 24 Jan 2019
21. Facebook, Quickstart for Graph API, 2019. <https://developers.facebook.com>. Accessed 24 Jan 2019
22. Freeman, M., Mcvittie, J., Sivak, I., Wu, J.: Viral information propagation in the Digg online social network. *Phys. A Stat. Mech. Appl.* **415**, 87–94 (2014)
23. Miritello, G., Moro, E., Lara, R.: Dynamical strength of social ties in information spreading. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **83**, 45102 (2011)
24. Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., Eguiluz, V.M.: Social features of online networks: the strength of intermediary ties in online social media. *PLoS ONE* **7**, e29358 (2012)
25. Xiong, F., Liu, Y., Zhang, Z.J., Zhu, J., Zhang, Y.: An information diffusion model based on retweeting mechanism for online social media. *Phys. Lett. A* **376**, 2103–2108 (2012)
26. Li, J.X., Tai, Z.Y., Zhang, R.C., Yu, W.R.: Online bursty event detection from microblog. In: Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE Computer Society, pp 865–870 (2014)
27. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003)
28. Li, Y., Chen, W., Wang, Y., Zhang, Z.L.: Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 657–666 (2011)
29. Nguyen, D.L., Nguyen, T.H., Do, T.H., Yoo, M.: Probability-based multi-hop diffusion method for influence maximization in social networks. *Wirel. Pers. Commun.* **93**, 903–916 (2017)
30. Tang, Y., Xiao, X., Shi, Y.: Influence maximization: near-optimal time complexity meets practical efficiency. *Eprint Arxiv*, pp. 75–86 (2014)
31. Chen, S., Fan, J., Li, G., Feng, J., Tan, K.L., Tang, J.: Online topic-aware influence maximization. In: Proceedings of the Vldb Endowment, vol. 8, pp. 666–677 (2015)

32. Domingos, P., Richardson, M.: Mining the network value of customers. In: Seventh Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, pp. 57–66 (2001)
33. Liu, L., Tang, J., Han, J., Jiang, M., Yang, S.: Mining topic-level influence in heterogeneous networks. In: ACM International Conference on Information & Knowledge Management, pp. 199–208 (2010)
34. Sun, P.G., Yang, Y.: Methods to find community based on edge centrality. *Phys. A Stat. Mech. Appl.* **392**, 1977–1988 (2013)
35. Campiteli, M.G., Holanda, A.J., Soares, L.D., Soles, P.R., Kinouchi, O.: Lobby index as a network centrality measure. *Phys. A Stat. Mech. Appl.* **392**, 5511–5515 (2013)
36. Sohn, J., Kang, D., Park, H., Joo, B.G., Chung, I.J.: An improved social network analysis method for social networks. In: *Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, Lecture Notes in Electrical Engineering, pp. 115–123. Springer, Amsterdam, The Netherlands (2014)
37. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. *J. Math. Soc.* **2**, 113–120 (1972)
38. Green, O., Bader, D.A.: Faster betweenness centrality based on data structure experimentation. In: *International Conference on Computational Science*, vol. 18, pp. 399–408 (2013)
39. Zhou, X., Chen, L.: Event detection over twitter social media streams. *VLDB J.* **23**, 381–400 (2014)
40. Wang, X., Zhai, C., Hu, X., Sproat, R.: Mining correlated bursty topic patterns from coordinated text streams. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 784–793 (2007)
41. AlSumait, L., Barbara, D., Domeniconi, C.: On-line LDA: adaptive topic models for mining text streams with applications to topic detection and tracking. In: *Proceedings of ICDM 2008: Eighth IEEE International Conference on Data Mining*, pp. 3–12 (2008)
42. Zhang, Y., Zhou, J., Cheng, J.: Preference-based top-K influential nodes mining in social networks. In: *Proceedings of the 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 1512–1518 (2011)
43. Zhou, D., Han, W., Wang, Y.: Identifying topic-sensitive influential spreaders in social networks. *Int. J. Hybrid Inf. Technol.* **8**, 409–422 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Lei-lei Shi received the B.S. degree from Nantong University, Nantong, China, in 2012, and the M.S. degree from Jiangsu University, Zhenjiang, China, in 2015. He is currently working towards the Ph.D. degree at the School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, China. His research interests include Event Detection, Data Mining, Social Computing and Cloud Computing.

Lu Liu is the Professor of Informatics and Head of Department of Informatics in the University of Leicester, UK. He received his Ph.D. degree from University of Surrey and M.S. degree from Brunel University. His research interests are in areas of Cloud Computing, Social Computing, Service oriented Computing and Peer-to-Peer Computing. He is the Fellow of British Computer Society (BCS).

Yan Wu received the M.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2009, and the Ph.D. degree from Tongji University, Shanghai, China, in 2014. He is currently a Lecturer with the School of Computer Science and Telecommunication Engineering in Jiangsu University, China. His research interests include formal methods, service-oriented Computing, and Cloud Computing.

Liang Jiang received the B.S. degree from Nanjing University of Posts and Telecommunications, China, in 2007, and the M.S. degree from Jiangsu University, Zhenjiang, China, in 2011. He is currently working towards the Ph.D. degree at the School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, China. His research interests include OSNs, Computer Networks and Network Security.

Ayodeji Ayorinde has a B.Sc. degree in Statistics (1992), M.Sc. degree in Statistics (1999) and M.Sc. degree in Epidemiology & Medical Statistics (2000), all from University of Ibadan, Nigeria. He has considerable industry experience in Data Mining, Analytics, Statistical Programming and Predictive Modelling. He is currently working towards his Ph.D. degree at the University of Derby, UK. Research interests includes Event Detection, Big Data Mining and Supervised Learning.

Affiliations

Lei-lei Shi¹ · Lu Liu² · Yan Wu¹ · Liang Jiang¹ · Ayodeji Ayorinde³

Lei-lei Shi
937109472@qq.com

Yan Wu
316119036@qq.com

Liang Jiang
13861358650@139.com

Ayodeji Ayorinde
A.Ayorinde@derby.ac.uk

- ¹ School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China
- ² School of Informatics, University of Leicester, Leicester, UK
- ³ School of Electronics, Computing and Mathematics, University of Derby, Derby, UK