Using Crowdsourced Data in Location-based Social Networks to Explore Influence Maximization

Ji Li Zhipeng Cai
Department of Computer Science
Georgia State University
Atlanta, Georgia 30302
Email: jli30@student.gsu.edu zcai@gsu.edu

Mingyuan Yan
Department of Computer Science
and Information Systems
University of North Georgia
Email: mingyuan.yan@ung.edu

Yingshu Li
Department of Computer Science
Georgia State University
Atlanta, Georgia 30302
Email: yili@gsu.edu

***Abstract*—Online social networks have gained significant popularity recently. The problem of influence maximization in online social networks has been extensively studied. However, in prior works, influence propagation in the physical world, which is also an indispensable factor, is not considered. The Location-Based Social Networks (LBSNs) are a special kind of online social networks in which people can share location-embedded information. In this paper, we make use of mobile crowdsourced data obtained from location-based social network services to study influence maximization in LBSNs. A novel network model and an influence propagation model taking influence propagation in both online social networks and the physical world into consideration are proposed. An event activation position selection problem is formalized and a corresponding solution is provided. The experimental results indicate that the proposed influence propagation model is meaningful and the activation position selection algorithm has high performance.**

I. Introduction

Nowadays, online social networks, such as Facebook and Twitter, have gained significant popularity. Online social networks have become a hot research topic in a variety of fields. There are plenty of existing works for online social networks focusing on topological characteristics, social group behavior, sentiment classification [1], [2], [3], [4], *etc*. Among all the research fields of online social networks, one of the most important research issues is the influence maximization problem. The influence maximization problem in an online social network is to find a subset of nodes named seed nodes, so that the spread of influence could be maximized if we use the seed nodes to spread the influence [5]. The influence maximization problem has been extensively studied in online social networks [6], [7], [8], [9], [10], [11]. However, the existing works have a common limitation that only the influence propagation in the online social network is considered. In fact, the influence of an event can also propagate through words of mouth among people in the physical world. Sometime, the influence by words of mouth propagation even has greater and faster impact compared with the influence of information in online social networks.

Based on the measurement results obtained from the actual datasets in Section III, we propose an influence propagation model for Location-Based Social Networks (LBSNs). An LBSN investigated in this paper is a type of online social networks in which people can share location-embedded information. It is able to bring social networks back to reality since the dimension of location helps bridge the gap between the physical world and online social networks. The influence propagation in both online social networks and the physical world is considered in our work. We are particularly interested in the problem of choosing an optimal position to activate an event so that the influence of the event can be maximized. Many problems in practice can be transformed to this problem, such as choosing an optimal position to hold an exhibition.

In this paper, we make use of mobile crowdsourced data obtained from location-based social network services to study the influence maximization problem. In summary, the main contributions of this paper are as follows.

1) Through analyzing the measurement results towards actual datasets, we identify an important factor of influence propagation which has been ignored by all the previous works in the literature. This factor motivates us to propose a novel network model and a novel influence propagation model. In the proposed model, propagation in both online social networks and the physical world is considered.
2) The proposed influence propagation model can be applied to a newly identified event activation position selection problem. A corresponding heuristic algorithm is then designed.
3) Extensive experimental results are presented, which demonstrate the meaningfulness of the proposed influence propagation model and also verify that the event activation position selection algorithm has high performance.

The rest of the paper is organized as follows. A novel network model is introduced in Section II. Selective measurement results towards two actual datasets are presented in Section III, based on which an influence propagation model is proposed in Section IV. Section V formalizes the event activation position selection problem and the corresponding solution is provided. Section VI discusses the extensive experimental results. Section VII presents the related works and Section VIII concludes the paper.

978-1-4673-9953-1/16/$31.00

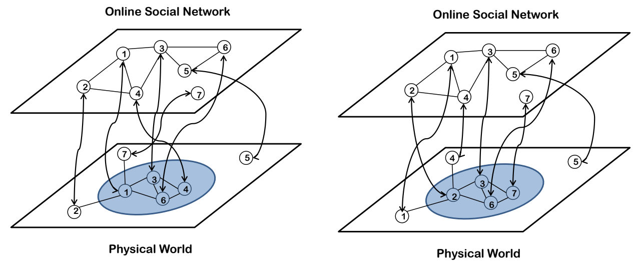
II. Network Model

Online social networks can be classified into undirected online social networks such as Facebook and directed online social networks such as Twitter. Statistics show that users behave differently in these two kinds of online social networks [12], [13]. In this paper, we focus on the influence propagation in undirected online social networks.

Suppose there are n users denoted by $V = \{v_1, v_2, \ldots, v_n\}$ deployed in region $[0,1)^2$. Each user is equipped with a smart phone and has an account in an undirected online social network. Each smart phone is equipped with a positioning component which can obtain its geographical position in the physical world and a mobile telecommunication component which is able to send information to the online social network. We use $v_i.x_t$ and $v_i.y_t$ to denote the x-coordinate and y-coordinate of user $v_i \in V$ at time t. For two users $v_i, v_j \in V$, $d(v_i, v_j, t) = \sqrt{(v_i.x_t - v_j.x_t)^2 + (v_i.y_t - v_j.y_t)^2}$ is used to denote the Euclidean distance between v_i and v_j at time t. Two users $v_i, v_j \in V$ are called neighbors in the physical world at time t if $d(v_i, v_j, t) \leq r_p$ where r_p is the predefined propagation radius. In this paper, we assume influence may propagate from an influenced user v_i to v_j if v_i and v_j are friends in the online social network or neighbors in the physical world. Then the online social network and users' graphical locations in the physical world can be described as a two-layer graph $G^t = (V, E_f, E_p^t)$ where V denotes the set of users. E_f and E_p^t are two undirected edge sets. For $\forall v_i, v_j \in V$, $(v_i, v_j) \in E_f$ if and only if v_i and v_j are friends in the online social network. Similarly, for time t and $\forall v_i, v_j \in V$, $(v_i, v_j) \in E_p^t$ if and only if $d(v_i, v_j, t) \leq r_p$. We assume E_f remains the same while E_p^t changes over time. In other words, a dynamic geometric graph is used to denote users' relationships in the physical world [14].

In this paper, we assume once an event such as an exhibition, advertisement or sports meeting is activated in the physical world, multiple users around the activation position will be influenced by this event. Influenced users may prorogate this event with their friends in online social networks or their neighbors in the physical world. That is, the influence of the event will propagate in both online social networks and the physical world simultaneously.

Fig.1 shows two instances, G^{t_1} and G^{t_2}, of our network model. Each two-layer graph has two sub-graphs to describe users' relationships in the online social network and in the physical world, respectively. Two nodes with the same ID denote the same user. Suppose an event is activated in the physical world and the users within the shadowed influencing region are influenced. Then the influence will propagate in both the online social network and the physical world. For G^{t_1} shown in Fig.1(a), we can see that although User 5 is not in the influencing region, User 5 may still be influenced since User 3 and User 6 are User 5's friends in the online social network and User 3 and User 6 are in the influencing region. Similarly, although User 1 and User 7 are not friends in the online social network, the influence may propagate from User 1 to User 7 since they are neighbors in the physical world.



(a) G^{t_1} (b) G^{t_2}

Fig. 1. Two instances of our network model.

Our first challenge is to develop an appropriate influence propagation model to simulate the influence propagation in the above circumstance. The influence propagation in both online social networks and the physical world should be considered in this model. Based on the proposed influence propagation model, we can then evaluate the influence of activating an event at different positions in the physical world. Obviously, the influence results may be quite different for different choices of event activation positions. Then our challenge is to seek an optimal position to activate an event so that the influence of this event can be maximized.

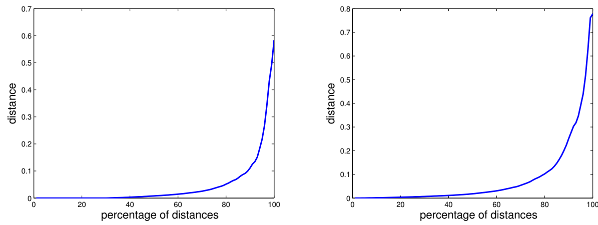
III. Measurements

To understand the characteristics of user behaviors in reality, two actual datasets named Brightkite and Gowalla from Stanford Large Network Dataset Collection are investigated in this paper [15]. Selected measurement results about users' positions in the physical world and their relationships in the online social networks are illustrated and analyzed. The inspirations on our propagation model construction and the activation position selection algorithm design are then obtained according to these measurement results.

A. Introduction to the Datasets

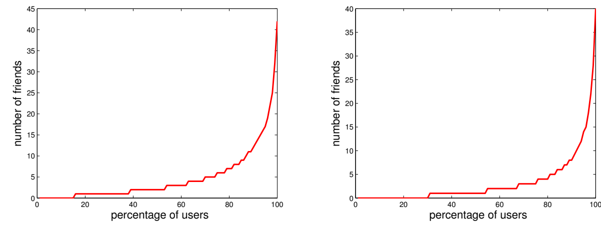
Both Brightkite and Gowalla are location-based social network service providers. Users can share their locations. When a user logs in, an individual record formatted as (*user-id, login-time, latitude, longitude, location-id*) is produced. It is difficult to detect users' movements when the distances of users movements are small and the scale of the physical world is large. To make sure the users' movements are correctly detected, we employ part of the original Brightkite and Gowalla datasets. In the employed two subsets, the users are distributed in $400km \times 400km$ rectangle regions. These regions include New York, Washington and Philadelphia where users are densely distributed. We use the random way point model to estimate users' positions [16]. The detailed position estimation method is omitted due to page limitation.

For the Brightkite dataset and Gowalla dataset, the investigated login records are over the period of April 2008 through October 2010 and December 2009 through October 2010, respectively. For simplicity, users' positions in the physical



(a) Brightkite (b) Gowalla

Fig. 2. The distribution of distances.



(a) Brightkite (b) Gowalla

Fig. 3. The distribution of numbers of friends.

world are standardized to $[0, 1)^2$ and the time stamps for the logins are standardized to $[0, 1)$. The details are summarized in Table I and Table II, respectively, where CC stands for connected component.

TABLE I
BRIGHTKITE DATASET

property	value	property	value
# users	3551	# logins	430657
# edges	9317	average login	121.278
average degree	5.248	# triangles	6738
# CC	569	average CC size	6.241
# nodes in largest CC	2907	# edges in largest CC	9228

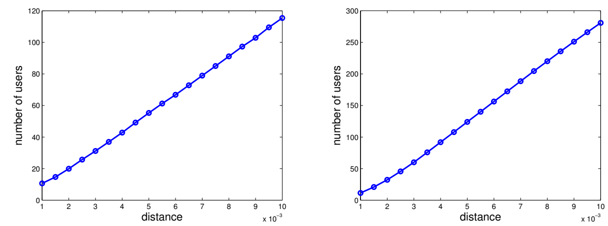
TABLE II
GOWALLA DATASET

property	value	property	value
# users	5231	# logins	297104
# edges	10134	average login	56.797
average degree	3.875	# triangles	11580
# CC	1778	average CC size	2.942
# nodes in largest CC	3114	# edges in largest CC	9676

B. Users' Positions and Number of Friends

We first measure the distribution of distances between users' different login positions. The results are shown in Fig.2. The results indicate that users' positions in the physical world have high stability from both the Brightkite dataset and the Gowalla dataset. In other words, users' positions in the physical world stay unchanged in many cases. For example, for the Brightkite dataset, 40% of the distances are less than 0.003059, which is about 1.2km.

The second measurement as shown in Fig.3 presents the distribution of the numbers of friends for different users in the online social networks. The results show that there is a strong skewness in this distribution for the two datasets. For example, the users in the Brightkite dataset who have the top 20% number of friends account for 75.16% of friends in the online social network. The rest 80% users only account for 24.84% of friends. That means the influence of different users in an online social network varies a lot. This phenomenon makes it possible that plenty of users get influenced in the end but only a few users get influenced initially. The experimental results in Fig.13 validate this presumption.



(a) Brightkite (b) Gowalla

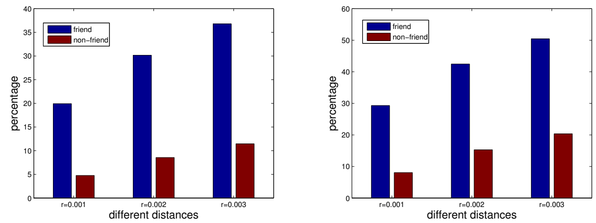
Fig. 4. The distribution of number of neighbors.

The third measurement is regarding the number of neighbors in the physical world within different distances. The results are presented in Fig.4, which indicate that there are many neighbors near a user in the physical world. For example, in the Gowalla dataset, there are about 60 users on average within the range of 0.003, which is about 1.2km. That means when a user tries to broadcast an event in the physical world, the user can easily find neighbors in the physical world and share the event information with them. That is, it is very convenient to propagate influence in the physical world.

C. Positions and Friendships

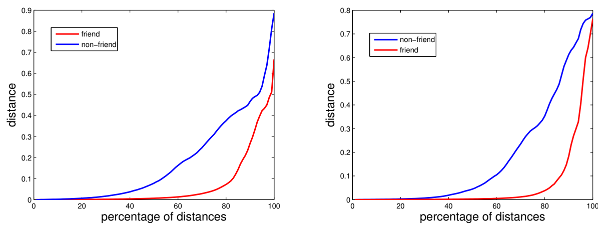
In this subsection, the results for both friends and non-friends are shown separately for comparison. Fig.5 shows the fraction of users within a given distance. We can see that in the Brightkite dataset, about 20% users keep the minimum distance less or equal to 0.001 to their friends (about 0.4km). However, for non-friends, only about 4.8% users' minimum distance is less or equal to 0.001. This indicates there is a high interdependency of geographical positions for friends in the online social networks. In other words, if users are friends with each other in the online social networks, their distances in the physical world tend to be closer. This phenomenon is also verified to be true by the Gowalla dataset. Therefore, it indicates that if two users are friends in an online social network, one may also be influenced by the other in the physical world.

Fig.6 presents the distribution of the minimum distance between users in the physical world. We can also see the high interdependency of geographical positions among friends. However, there do exist friends in online social networks who are far away from each other in the physical world in some cases. For example, in the Brightkite dataset, around 10%



(a) Brightkite (b) Gowalla

Fig. 5. The percentage of users within a given distance.

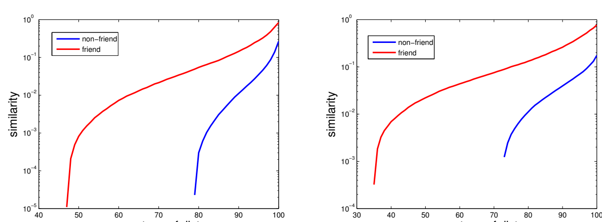


(a) Brightkite (b) Gowalla

Fig. 6. The distribution of minimum distances.

of the minimum distances between friends are greater than 0.267084(about 107km). That means, with the help of online social networks, it is possible that the influence of an event can spread widely in the physical world. Furthermore, for non-friends, the minimum distances between them are quite small in some cases. For example, in the Gowalla dataset, about 18% of the minimum distances between non-friends are less than 0.002 (about 0.8km). That means it is possible for the influence to spread between two users in the physical world even if they are not friends in online social networks.

The next measurement is on trajectory similarity in the physical world as shown in Fig.7. We create a trajectory vector for each user by setting the i-th component to be the number of times that a user checked-in at location i. The trajectory similarity of two users is defined as the cosine similarity between their trajectory vectors [17]. The results demonstrate that friends in online social networks have higher trajectory similarity in the physical world. For example, in the Brightkite dataset, the trajectory similarity is greater than 0.01 for about 40% friends. However, for non-friends, this ratio is only about 10%.



(a) Brightkite (b) Gowalla

Fig. 7. The distribution of trajectory similarities.

IV. Influence Propagation Model

Based on the measurement results obtained in Section III, an influence propagation model is proposed in this section. In this model, propagation in both online social networks and the physical world is considered. The phenomenon named *cross propagation* is introduced in the last subsection. In order to describe the influence propagation model, some attributes for events and users are employed, as summarized in Table III.

TABLE III
RELATIVE ATTRIBUTES

Attribute	Description
$E.type$	Type of event
$E.x_0$	X-coordinate of activation position
$E.y_0$	Y-coordinate of activation position
$E.t_0$	Activation time
$E.r_0$	Influence radius
$E.t_{init\_pro}$	Initial propagation time
$E.t_{add\_pro}$	Additional propagation time
$v.interest$	Interest of user v
$v.x_t$	X-coordinate of user v's position at time t
$v.y_t$	Y-coordinate of user v's position at time t
$v.share\_phy$	Set of time for v to share the event

An event has one or multiple types, such as *sport*, *music*, *exhibition* and so on. The set $E.type$ is used to denote the type of event E. Each user v has one or multiple interests, such as *music*, *art* and so on. The set $v.interest$ is used to denote the interests of user v. We assume $default\_type \in E.type$ for any event E and $default\_type \in v.interest$ for any user v.

A. Basic Influence Propagation Model

During the influence propagation period, multiple factors such as users' interest may affect the user's probability to be influenced. We use function $I(x, I_{max}) = (I_{max} - 1)\sqrt{1-(1-x)^2}+1$ to measure the increase of influencing probability, where x is the input probability increasing parameter and I_{max} is the maximum increase. Then we have $I(0, I_{max}) = 1$, $I(1, I_{max}) = I_{max}$, $\frac{\partial I(x,I_{max})}{\partial x} > 0$ and $\frac{\partial^2 I(x,I_{max})}{\partial x^2} < 0$ for $x \in (0,1)$.

Obviously, if the type of an event closely matches user v's interest, it is more likely that v can be influenced by this event and then further propagates the event. We use $I_1(E, v)$ to measure the increase of influence and share probability based on the match between user's interest and the event type. We define $I_1(E, v)$ as

$$I_1(E, v) = I(J(E.type, v.interest), I_{max1})$$

where I_{max1} is the upper bound of increase which is a predefined constant. $J(E.type, v.interest)$ is the the Jaccard similarity coefficient between $E.type$ and $v.interest$ which is defined as

$$J(E.type, v.interest) = \frac{|E.type \cap v.interest|}{|E.type \cup v.interest|}.$$

For an event E, $E.x_0$ and $E.y_0$ are used to denote the x-coordinate and y-coordinate of the activation position of E. $E.t_0$ is the time when E is activated. An event has two other attributes, which are the initial propagation time $E.t_{init\_pro}$ and the additional propagation time $E.t_{add\_pro}$. Once event E is activated, it is propagated in the physical world during the initial propagation period $[E.t_0, E.t_0 + E.t_{init\_pro})$. Some users around the activation position of E will be influenced. After the initial influence propagation period, the event will be propagated in online social networks and the physical world simultaneously during the additional propagation period $[E.t_0 + E.t_{init\_pro},\ E.t_0 + E.t_{init\_pro} + E.t_{add\_pro})$. Each propagation will be introduced in details in the rest of this section.

B. Initial Influence Propagation

Let $R(x, y, r)$ denote the circular region around position (x, y) with radius r which is defined as

$$R(x, y, r) = \{(x', y') \in [0, 1)^2 | (x' - x)^2 + (y' - y)^2 \leq r^2\}.$$

For $\forall v \in V$, $v.x_t$ and $v.y_t$ are the x-coordinate and y-coordinate of v's position at time t. Then we define set $S(x, y, r, t)$ as

$$S(x, y, r, t) = \{v \in V | (v.x_t, v.y_t) \in R(x, y, r)\}$$

which includes all the users in $R(x, y, r)$ at time t. Based on this definition, we define set $S(x, y, r, t_1, t_2)$ as

$$S(x, y, r, t_1, t_2) = \{v | \exists v \in S(x, y, r, t), t \in [t_1, t_2)\}$$

which includes all the users who have appeared in $R(x, y, r)$ during the time period $[t_1, t_2)$.

The influence radius of event E is denoted by $E.r_0$. Once E is activated, $\forall v \in S(E.x_0, E.y_0, E.r_0, E.t_0, E.t_0 + E.t_{init\_pro})$ will be influenced by E with probability $p(E, v, init\_inf) = \min(p_1 I_1(E, v) I_2(E, v), 1)$ where p_1 is the base influence probability which is a predefined constant. $I_1(E, v)$ is introduced in Section IV-A. $I_2(E, v)$ is used to denote the increase of the influencing probability which depends on the time for v to stay in $R(E.x_0, E.y_0, E.r_0)$. $I_2(E, v)$ is defined as

$$I_2(E, v) = I(\frac{T(E, v)}{E.t_{init\_pro}}, I_{max2})$$

where I_{max2} is the upper bound of increase which is a predefined constant. $T(E, v)$ is used to denote the time v stays in $R(E.x_0, E.y_0, E.r_0)$ during period $[E.t_0, E.t_0 + E.t_{init\_pro})$.

C. Influence Propagation in Online Social Networks

Let v be a user who has already been influenced by an event. After v is influenced, once v logs in an online social network, v shares the event information with v's friends with probability

$$p(v, osn\_share) = \min(p_2 I_1(E, v), 1)$$

where p_2 is the predefined base sharing probability.

Assume as v's friend, user u receives the event information when u logs into the online social network at time t. Since an event may be shared by multiple friends of u, u may see multiple copies of the event information. We use $n_r(E, u, t)$ to denote the number of descriptions of event E received by user u at time t. When u logs into the online social network at time t, u will not be influenced if $n_r(E, u, t) = 0$. Otherwise, u may be influenced by E. The probability that u will be influenced by event E can be defined as

$$p(E, u, osn\_inf) = \min(p_3 I_1(E, u) I_3(u, t), 1)$$

where p_3 is a predefined base influencing probability. $I_3(u, t)$ denotes the increase of influencing probability based on the number of received copies of E's descriptions which is defined as

$$I_3(u, t) = I(\min(\frac{n_r(E, u, t) - 1}{n_{max}}, 1), I_{max3})$$

where I_{max3} is the upper bound of increase which is a predefined constant. n_{max} is used to standardize $n_r(E, u, t)$ which is a predefined constant.

D. Influence Propagation in the Physical World

If user v is influenced by event E, v will continue to propagate the influence of E within the physical world with probability

$$p(v, pw\_share) = \min(p_4 I_1(E, v), 1).$$

If user v decides to propagate event E within the physical world, $v.share\_phy$ denotes the set of different time instances for v to share E in the physical world. For $\forall t \in v.share\_phy$, v will choose a user $u \in S(v.x_t, v.y_t, r_p, t)$ and share E with u if $S(v.x_t, v.y_t, r_p, t) \neq \emptyset$. $v.share\_phy$ is randomly generated in our model since it is almost impossible to acquire $v.share\_phy$ in practice.

Obviously, during the propagation of E in the physical world, v tends to choose one of its friends u in the online social network in $S(v.x_t, v.y_t, r_p, t)$ to share E. On the other hand, if user v shares E with user u, u will more likely to be influenced by E if v is u's friend in the online social network. Then $I_4(v, u)$ is used to offer different weights for users in an online social network. $I_4(v, u)$ is defined as

$$I_4(v, u) = \begin{cases} c & \text{if } (v, u) \in E_f \\ 1 & \text{otherwise} \end{cases}$$

where $c > 1$ is a predefined constant. For $\forall t \in v.share\_phy$, user v will choose user $u \in S(v.x_t, v.y_t, r_p, t)$ and share event E with u. The roulette wheel method is used to decide which user u to be chosen [18]. $I_4(v, u)$ is used to calculate the weight in the roulette wheel method. If v shares E with u, u will be influenced by E with probability

$$p(E, u, phy\_inf) = \min(p_5 I_1(E, u) I_4(v, u), 1)$$

where p_5 is a predefined probability and $I_1(E, v)$ is introduced in Section IV-A.

E. Cross Propagation

During the additional propagation period $[E.t_0 + E.t_{init\_pro}, E.t_0 + E.t_{init\_pro} + E.t_{add\_pro})$, the influence of E will propagate in online social networks and the physical world simultaneously using the model introduced in Section IV-C and Section IV-D. Influence does not simply propagate in online social networks or the physical world separately. Such a phenomenon is called *cross propagation* with the following two properties.

1) Influence may propagate from online social networks to the physical world. For example, once user v is influenced by event E in an online social network, v could share E with v's neighbors in the physical world. Then v's neighbors in the physical world may be influenced by E even if some of the influenced neighbors are not v's friends in the online social network.
2) Influence may propagate from the physical world to online social networks. For example, once user v is influenced by event E through words of mouth from v's neighbors in the physical world, v may share E with v's friends in an online social network. Then v's friends in the online social network could be influenced by E.

Cross propagation between the online social network and the physical world is continuous during the additional propagation period $[E.t_0 + E.t_{init\_pro}, E.t_0 + E.t_{init\_pro} + E.t_{add\_pro})$. Cross propagation truthfully reflects the influence propagation in reality.

V. Optimal Event Activation Position Selection

We formalize the event activation position selection problem followed by the proposed solution in this section.

A. Problem Description

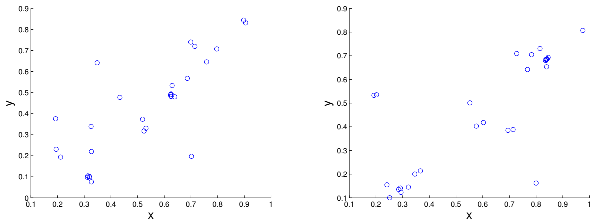
In this paper, we assume the initial propagation time, the additional propagation time and the influence radius of an event are fixed. The activation position of an event is chosen from a candidate position set $\mathcal{C}$. pos is used to denote a position in $\mathcal{C}$. We use $F(pos)$ to denote the number of final influenced users if we activate event E in physical position pos. The influence propagation model in Section IV is used to calculate the value of $F(pos)$. Obviously, since the distribution of users in the physical world is not uniform, even for the same event, the numbers of final influenced people may differ a lot due to different choices of event activation position. Therefore, we propose the activation position selection problem as a problem to select the best position from candidate position set $\mathcal{C}$ to activate the event so that its influence can be maximized. The problem can be formalized as follows.

Input: candidate position set $\mathcal{C}$, event E, two-layer graph $G^t = (V, E_f, E_p^t)$.

Output: $\arg \max_{pos \in \mathcal{C}} F(pos)$.

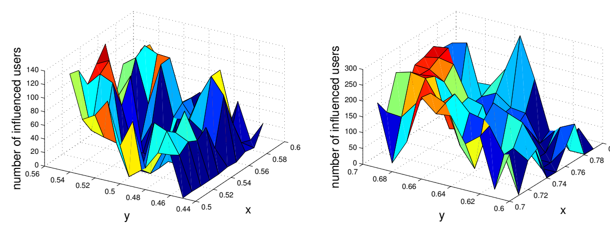
B. Solution to the Activation Position Selection Problem

Some previous works such as [5] and [19] proposed algorithms to find a small set of seed nodes in a social network



(a) Brightkite (b) Gowalla

Fig. 8. User distribution.



(a) Brightkite (b) Gowalla

Fig. 9. Function $F(pos)$'s values.

to maximize the spread of influence. But these algorithms cannot be used in our work directly since the initial influencing radius of an event is limited while the seed nodes' distribution may be highly dispersive in the physical world. Suppose we adopt the independent cascade model and greedy algorithm in [19] to calculate the top-30 influencing users in an online social network. The results are shown in Fig.8. We can see the distributions of these users are highly dispersive when the event is activated. It is almost impossible to find a suitable event activation position so that all of these users are within the initial influence region.

A straightforward method to solve this problem is to use mathematical methods, such as considering the monotonicity of function $F(pos)$. However, as Fig.9 shows, for some candidate position set $\mathcal{C}$ such as $\{(x, y) \mid 0.5 \leq x \leq 0.6, 0.45 \leq y \leq 0.55\}$ for the Brightkite dataset or $\{(x, y) \mid 0.7 \leq x \leq 0.8, 0.6 \leq y \leq 0.7\}$ for the Gowalla dataset, the values of function $F(pos)$ distribute highly irregularly. Mathematical methods do not work well in this case. Another possible solution is to calculate the number of influenced people for each position in the candidate position set $\mathcal{C}$. Then the position which maximizes the number of influenced people is chosen. This method does not fit for the circumstance that the candidate position set $\mathcal{C}$ is an infinite set. Finally, we have to propose an algorithm to seek a proper activation position for an event in a heuristic way instead. The basic idea is as follows.

1) Randomly select N_s positions in the candidate position set $\mathcal{C}$.
2) Select a position among these N_s positions using the roulette wheel method [18]. Replace the selected position with a new position if the new position has greater F value. This process is carried out for N_i times.

3) Return the position with the largest F value.

The following method is used to calculate the new position pos_2 from the selected position pos_1. dis is the Euclidean distance between the two positions.

1) $N_c[pos_1]$ records how many times pos_1 is selected. Δ_{init} and α ($\alpha > 1$) are predefined constants.
2) Calculate the upper bound of $dis(pos_1, pos_2)$ by $\Delta = \frac{\Delta_{init}}{\alpha^{N_c[pos_1]}}$.
3) Randomly select $pos_2 \in \{pos \in \mathcal{C} \mid dis(pos, pos_1) \leq \Delta\}$.

The detailed algorithm is shown in Algorithm 1. $rand$ denotes a random value in $[0, 1)$.

Algorithm 1 Optimal Activation Position Selection Algorithm

Input: candidate position set $\mathcal{C}$, event E, two-layer graph $G^t = (V, E_f, E_p^t)$

Output: position to activate event E

```
for i = 1 to N_s do
    P[i]= randomly selected element in C;
end for
for i = 1 to N_s do
    N_c[i] = 0;
end for
for i = 1 to N_i do
    find the minimum j satisfying (Σ_{k=1}^{j} F(P[k])) / (Σ_{k=1}^{N_s} F(P[k])) > rand;
    Δ = Δ_init / α^{N_c[j]};
    randomly select pos' ∈ {pos ∈ C | dis(pos, P[j]) ≤ Δ};
    if F(pos') > F(P[j]) then
        P[j] = pos';
    end if
    N_c[j]++;
end for
return arg max_{pos∈P} F(pos);
```

However, due to the high complexity of the influence propagation model, the calculation of function F is slow. That means Algorithm 1 has high computation cost. Then we propose another objective function F' which also considers influence propagation in both online social networks and the physical world. F' is defined as

$$F'(pos) = \sum_{u \in U} |u.friends| + \frac{\sum_{v \in Neg(u,t)} |v.friends|}{|Neg(u,t)|}$$

where the symbols in F' are introduced in Table IV. Then we have the F'-based heuristic algorithm. Its process is the same as Algorithm 1. The only difference is to use F' as the objective function rather than F to evaluate the event activation position.

VI. Experimental Results

In this section, we evaluate the proposed influence propagation model and the activation position selection algorithm through experiments. The parameters are listed in Table V. The candidate position set $\mathcal{C}$ for the activation position selection algorithm is shown in Table VI.

TABLE IV
Symbols in Function F'

t	a randomly selected time instance in $[E.t_0 + E.t_{init\_pro},\ E.t_0 + E.t_{init\_pro} + E.t_{add\_pro})$
$v.friends$	the set of user v's friends in the online social network
$pos.x$ / $pos.y$	the x-coordinate / y-coordinate of position pos
$Neg(u,t)$	u's neighbors in the physical world at time t
U	$S(pos.x, pos.y, E.r_0, E.t_0)$

TABLE V
Parameters for the Experiments

parameter	value	parameter	value	parameter	value
$E.r_0$	0.01	$E.t_0$	0.5	$E.t_{init\_pro}$	0.02
$E.t_{add\_pro}$	0.2	I_{max1}	3	I_{max2}	1.5
I_{max3}	6	r_p	0.01	c	5
n_{max}	10	α	2	Δ	0.1
N_s	10	N_i	10		

TABLE VI
Candidate Position Set

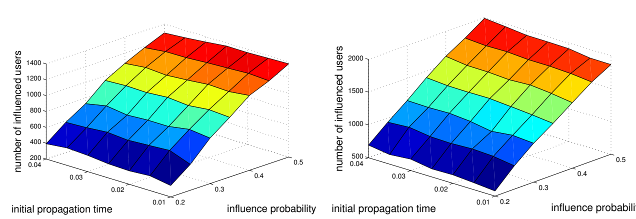
dataset	candidate position set
Brightkite	$\{(x, y) \mid 0.5 \leq x \leq 0.6, 0.45 \leq y \leq 0.55\}$
Gowalla	$\{(x, y) \mid 0.7 \leq x \leq 0.8, 0.6 \leq y \leq 0.7\}$

A. Number of Influenced Users

The first group of experiments are about the number of influenced users. We set different influencing probabilities in these experiments. For a given influencing probability p, we set $p_1 = p$ and $p_2 = p_3 = p_4 = p_5 = 1.5p$.

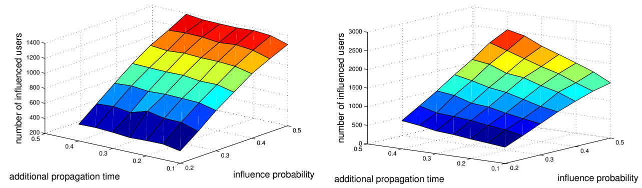
The first experiment is about the relationship among the initial propagation time, influencing probability and the number of influenced users. The results are shown in Fig.10. We can see that the number of influenced users increases with the increase of influencing probability, while the number of influenced users almost remains the same with the increase of the initial propagation time. A reason for this phenomenon is the high stability of users' traces shown in Fig.2. Another reason is that we use $\frac{T(E,v)}{E.t_{init\_pro}}$ instead of $T(E, v)$ itself as a parameter of $I_2(E, v)$.

The second experiment is about the relationship among the additional propagation time, influencing probability and the number of influenced users. The results are presented in Fig.11. Similarly, we can see that the number of influenced users increases with the increase of influencing probability. In general, the number of influenced users increases with the



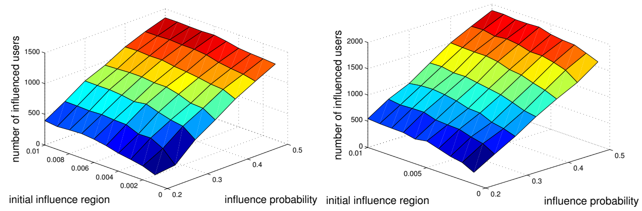
(a) Brightkite (b) Gowalla

Fig. 10. The number of influenced users for different initial propagation time.



(a) Brightkite (b) Gowalla

Fig. 11. The number of influenced users for different additional propagation time.



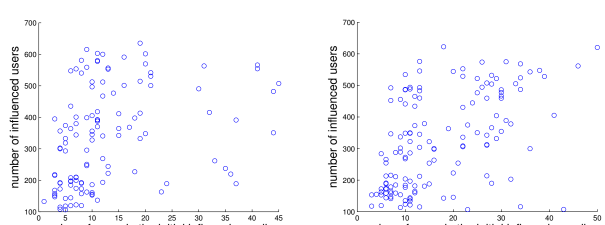
(a) Brightkite (b) Gowalla

Fig. 12. The number of influenced users for different initial influence radius.

increase of the additional propagation time since there will be more sharing of event in both online social networks and the physical world. This trend is especially obvious in the Gowalla dataset as shown in Fig.11(b).

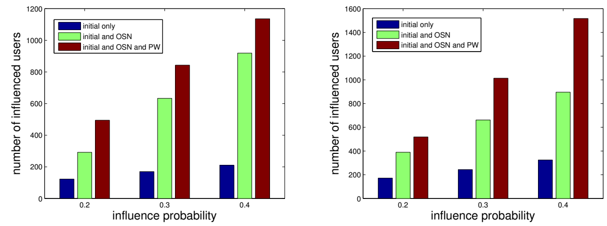
The third experiment is to find the relationship among the initial influence radius, influencing probability and the number of influenced users. Fig.12 shows the result. It can be seen that the number of influenced users increases with the increase of the initial influence radius. The number of influenced users increases with a faster speed in the beginning but slower afterwards since users are densely deployed in a small region.

The fourth experiment is about the relationship between the number of users within the initial influencing region and the number of influenced users. The results are shown in Fig.13. We can see that there is little interdependency between the number of users in the initial influencing region and the number of influenced users. So when we try to determine the event activation position to maximize influence, we cannot simply choose the position with the maximum number of users.



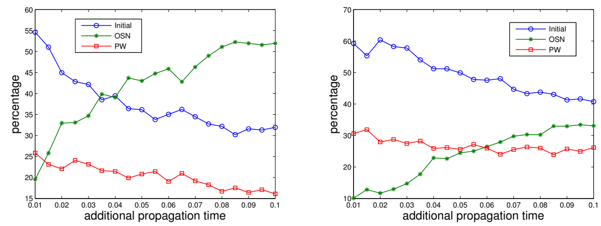
(a) Brightkite (b) Gowalla

Fig. 13. The number of influenced users *vs*. the number of users in the initial influencing region.



(a) Brightkite (b) Gowalla

Fig. 14. Comparison of different propagation models.



(a) Brightkite (b) Gowalla

Fig. 15. Comparison of different influence manners.

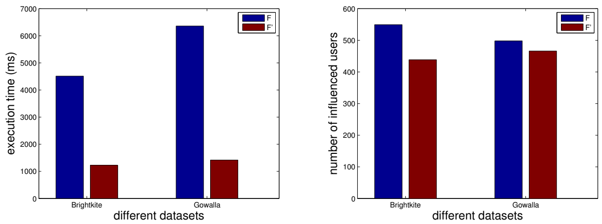
B. Distribution of Influenced Users

The distribution of influenced users are investigated in this subsection. The results are shown in Fig.14 and Fig.15. We first want to compare the number of influenced users under three different propagation models with different influencing probabilities. The first model only considers the process of initial influence propagation. The second model includes both the initial propagation and the propagation in online social networks. The third model considers initial propagation, propagation in online social networks and propagation in the physical world. From Fig.14, we can see that the third model has the largest number of influenced users since it takes the propagation in both online social networks and the physical world into consideration.

The next experiment is to compare the ratios of the users influenced by different manners. We set p_1=0.2 and $p_2 = p_3 = p_4 = p_5 = 0.3$. As can be seen in Fig.15, for both the Brightkite and Gowalla datasets, the ratio of users influenced in online social networks increases with the increase of additional propagation time. The ratio of the users influenced in the physical world almost remains the same. While the ratio of users influenced during the initial influence propagation decreases with the increase of the additional propagation time since the number of these users almost remains the same but more and more users are influenced rapidly.

C. The Optimal Activation Position Selection Algorithm

We compare the F-based heuristic algorithm and F'-based heuristic algorithm on the aspects of execution time and the number of influenced users by activating the event at the output position. The results are shown in Fig.16. We can see that the F'-based heuristic algorithm has much shorter execution time on both the Brightkite and Gowalla datasets as expected.



(a) execution time (b) number of influenced users

Fig. 16. Comparison between F-based and F'-based heuristic algorithms.

However, because F' cannot perfectly evaluate the influence propagation, activating the event at the position derived by the F'-based heuristic algorithm results in fewer influenced users.

VII. Related Works

Recently, Location-Based Social Networks (LBSNs) have gained significant popularity. The prevalence of LBSNs contributes massive data with users' location information embedded. Many existing works have studied LBSNs from multiple aspects [20], [21], [22], [23], [24].

The works in [20] and [21] propose multiple influence maximization algorithms for LBSNs. However, their focus is not to select an event activation position but to select top influencing users. Then their results cannot be directly employed for our work. The work in [17] uses the Brightkite and Gowalla datasets to study friendships in online social networks and users' movements in the physical world. The work in [22] proposes a friendship prediction approach by fusing the topology and geographical features in LBSNs. Based on the related information exposed in LBSNs, the work in [23] analyzes the social-spatial influence and social-temporal influence. A probabilistic model of user mobile behaviors is proposed. The work in [24] studies the impact of social relations hidden in LBSNs. A new social influence-based user recommendation framework is proposed.

VIII. Conclusion

In this paper, a new event influence propagation model is proposed based on the measurement results of two actual datasets. This model is motivated by the fact that influence propagates in both online social networks and the physical world, which has been overlooked by all the previous works. We also propose a new network model, based on which an event activation position selection problem is defined. A corresponding heuristic algorithm is designed. The experimental results validate the meaningfulness of the influence propagation model and show that the proposed activation position selection algorithm has high performance.

Acknowledgment

This work is partly supported by the NSF under grant No.CNS-1252292, and the NSFC under grants No.61370084, No. 61373083 and No.61502116.

References

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *ACM WWW*, 2007, pp. 835–844.

[2] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook friends: Social capital and college students use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[3] T. Zagibalov and J. Carroll, "Automatic seed word selection for unsupervised sentiment classification of chinese text," in *Proceedings of the 22nd International Conference on Computational Linguistics*, 2008, pp. 1073–1080.

[4] Z. He, Z. Cai, and X. Wang, "Modeling propagation dynamics and optimal countermeasures of the social network rumors," in *IEEE ICDCS*, 2015, pp. 205–214.

[5] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD*, 2003, pp. 137–146.

[6] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *ACM SIGKDD*, 2009, pp. 199–208.

[7] J. Li and Y. Yu, "Scalable influence maximization in social networks using the community discovery algorithm," in *IEEE ICGEC*, 2012, pp. 284–287.

[8] M. Han, M. Yan, Z. Cai, and Y. Li, "An exploration of broader influence maximization in timeliness networks with opportunistic selection," *Journal of Network and Computer Applications*, 2016.

[9] T. Shi, J. Wan, S. Cheng, Z. Cai, Y. Li, and J. Li, "Time-bounded positive influence in social networks," in *IIKI*, 2015.

[10] X. Wang, Y. Lin, Y. Zhao, L. Zhang, Y. Liang, and Z. Cai, "A novel approach for inhibiting misinformation propagation in human mobile opportunistic networks," *Peer-to-Peer Networking and Applications*, In Press, 2016.

[11] X. Wang, Y. Lin, L. Zhang, and Z. Cai, "A double pulse control strategy for misinformation propagation in human mobile opportunistic networks," in *WASA*, 2015, pp. 571–580.

[12] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using facebook. com," *Social networks*, vol. 30, no. 4, pp. 330–342, 2008.

[13] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *ACM WWW*, 2010, pp. 591–600.

[14] J. Pach, "Geometric graph theory," Cambridge University Press, Tech. Rep., 1999.

[15] J. Leskovec and A. Krevl, "Snap datasets: Stanford large network dataset collection," Jun. 2014.

[16] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile computing*. Springer, 1996, pp. 153–181.

[17] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *ACM SIGKDD*, 2011, pp. 1082–1090.

[18] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 6, pp. 2193–2196, 2012.

[19] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, 2012.

[20] Y.-T. Wen, P.-R. Lei, W.-C. Peng, and X.-F. Zhou, "Exploring social influence on location-based social networks," in *IEEE ICDM*, 2014, pp. 1043–1048.

[21] T. Zhou, J. Cao, B. Liu, S. Xu, Z. Zhu, and J. Luo, "Location-based influence maximization in social networks," in *ACM CIKM*, 2015, pp. 1211–1220.

[22] H. Luo, B. Guo, Z. W. Yu, Z. Wang, and Y. Feng, "Friendship prediction based on the fusion of topology and geographical features in lbsn," in *IEEE HPCC_EUC*, 2013, pp. 2224–2230.

[23] Y. Song, Z. Hu, X. Leng, H. Tian, K. Yang, and X. Ke, "Friendship influence on mobile behavior of location based social network users," *Journal of Communications and Networks*, vol. 17, no. 2, pp. 126–132, 2015.

[24] Y.-T. Wen, P.-R. Lei, W.-C. Peng, and X.-F. Zhou, "Exploring social influence on location-based social networks," in *IEEE ICDM*, 2014, pp. 1043–1048.