Exploiting Mobility for Location Promotion in Location-based Social Networks

Wen-Yuan Zhu
National Chiao Tung University
Hsinchu, Taiwan
wyzhu@cs.nctu.edu.tw

Wen-Chih Peng
National Chiao Tung University
Hsinchu, Taiwan
wcpeng@cs.nctu.edu.tw

Ling-Jyh Chen
Academia Sinica
Taipei, Taiwan
cclljj@iis.sinica.edu.tw

***Abstract*—In this paper, we target the location promotion problem in location-based social networks (LBSNs). The location promotion problem is given a location, we select a set of users as seeds to influence as many users as possible who are likely to visit a selected location. Specifically, we model the location promotion problem as an influence maximization problem on a graph and explore the independent cascading diffusion model on the graph. To determine the propagation probability of the edges of our proposed graph, the relation between users and the selected location should be detected. A property of LBSN is that the major reason of users visiting a location is based on their mobility. Therefore, we propose a mobility model DMM (Distance-based Mobility Model) to model each user's mobility. DMM exploits random walk with restart and the power law property of users' movements. Based on DMM and the selected location, the propagation probability of edges can be derived. In the evaluation, we show the performance of our proposed algorithms on two real datasets.**

I. Introduction

Location promotion reveals when a location needs to attract many visitors, such as shops, hotels and event locales. A traditional way is to advertise to promote a specific location on different media. However, the cost of buying media advertising is huge, and it is not effective since most of viewers are not interested in it. Some service providers have noticed the increment in mobile devices. Thus, some companies provide recommendation services on mobile phones via SMS or applications on smart phones, such as Kellton Tech[1]. However, it is hard to convince users to install an advertisement application on their mobile phones unless there are other benefits. Nevertheless, with the increment of location-based services, especially in location-based social networks (LBSNs), such as Foursquare[2], users use the "check-in" service to share their current location with their friends for more social interactions in LBSNs. Therefore, some LBSN providers are beginning to provide location promotion services for businesses on their platforms[3].

The location promotion problem in LBSNs is to select a fixed number of seeds to maximize number of potential users who will visit the selected location influenced by the seeds. Figure 1 shows an example of an LBSN.First, we select the location ℓ_1, which is the target of promotion, and then we aim to select k users as seeds to whom send information about the selected location ℓ_1. Their friends will receive the information via social connections in LBSN so as to attract as many users as possible to visit the selected location ℓ_1 via social connections.

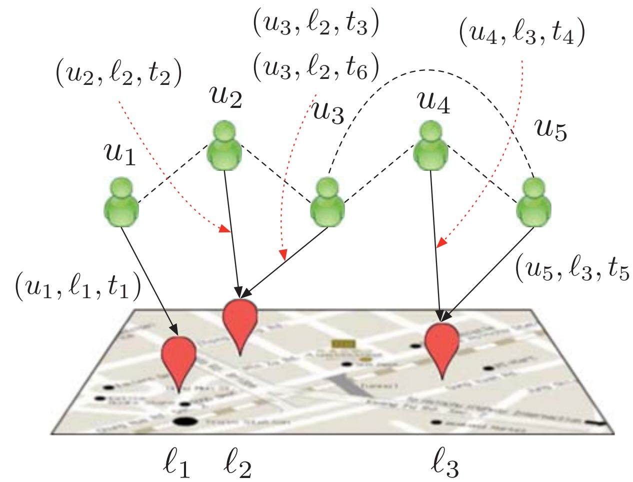


Fig. 1. An example of the location promotion problem in LBSNs

The location promotion problem in LBSNs is similar to influence maximization on social networks. However, there are three unique features in the location promotion problem. First, an LBSN consists of not only social connections but also the check-in history of users, whereas traditional social networks only consist of social connections. Second, location promotion has to select a specific location to promote, while influence maximization does not. Third, the influence consists of not only social connection but also the movement behavior of users in LBSNs [6][23]. For example, user u_3 and u_5 share a social connection. However, user u_3 prefers location ℓ_2, and user u_5 prefers location ℓ_3. Moreover, the distance between ℓ_2 and ℓ_3 is long. Therefore, it is not suitable to promote location ℓ_2 for user u_5 using this social connection since their movement behavior is different. Due to the three unique features, existing algorithms for influence maximization on social networks can not directly apply to this problem.

In this paper, we utilize the independent cascading model (ICM) to describe information influence in LBSNs. Moreover, to determine the propagation probability of the edges of the graph, we exploit the feature of LBSNs that the major reason for users to visit a location is based on their mobility [23].

[1] http://www.kelltontech.com/
[2] http://foursquare.com/
[3] http://business.foursquare.com/

Based on this feature, the propagation probability of edges is related to the probability of users visiting the selected location. To evaluate whether a user visits the selected location, we have to infer users' mobility using their check-in history in LBSN. To archive this problem, spatial and temporal sparsity should be dealt with in check-in history on the LBSN. Traditionally, to capture movement behavior, sequential pattern mining methods can be utilized to extract frequent sequential patterns from trajectories [10][14]. However, due to the spatial and temporal sparsity, it is hard to extract frequent sequential patterns from the check-in history in LBSNs. In [6], the authors proposed two mobility models, PMM and PSMM, to model each user's mobility in different time. PSMM consists of three Normal distributions, which represent spacial, temporal and social, respectively. However, it is hard to use PSMM to determine the propagation probability of each edge. Since the probability of one location in a two-dimensional space is zero, it has to select a scope around the selected location to calculate the probability of user visiting the selected location from probability density function. Thus, it is hard to decide the scope of the selected location. Moreover, PSMM needs to have mobility models in different time slots, but it is hard to collect enough data in LBSN to train mobility model in different time slots. Therefore, we propose a distance-based mobility model DMM (Distance-based Mobility Model), which consists of two modules. First, we utilize random walk with restart to estimate probable locations with their probability of being stayed in [5]. Furthermore, due to the space sparsity of check-in history, the number of probable locations may be few. Thus, we observe and exploit the power law property of the movement distance to estimate the probability of the user moving from one location to the selected location. Based on DMM, it is easy to derive the probability of users visiting the selected location for the propagation probability. As such, we can utilize methods of the influence maximization problem to find k seeds with the purpose of maximizing the influence spread to attract as many as users possible to visit the selected location.

In summary, our major contributions are outlined as follows:

- We first model the location promotion problem in LBSNs as an influence maximization problem on ICM in a graph.
- To address the location promotion problem, we develop a graph structure to represent LBSN and then further explore independent cascading model on the LBSN graph.
- We propose a mobility model DMM to capture users' mobility using random walk with restart and the power law property of the movement distance.
- We have conducted comprehensive experiments on two real datasets, and the experimental results show that the proposed algorithms are able to promote locations in LBSNs effectively.

The remainder of this paper is organized as follows: Section II discusses several related works. Section III gives a formal definition for the location promotion problem and shows how to model location promotion problem as an influence maximization problem. Section IV presents our proposed mobility model DMM to capture each user's mobility in LBSN. The experimental results are shown in Section V. Section VI concludes this paper.

II. Related Work

In this paper, the goal is to promote a specific location in LBSNs. Our work is related to the information diffusion models and influence maximization algorithms on social networks. Thus, we will present some existing work about diffusion models and influence maximization algorithms. On the other hand, due to the features of LBSNs, we have to consider the movement behavior in our problem. Furthermore, we will describe the existing works on capturing movement behavior from trajectories.

A. Capturing Movement Behavior

To capture movement behavior from trajectories, there are already many existing works for this problem. Since the trajectory is a sequence of locations ordered by time, many works focus on mining sequential patterns from trajectories. Moreover, trajectories carry not only temporal but also spatial information. Therefore, spatio-temporal patterns can be driven from trajectories to represent a user's movement behavior. In [3], the authors use segments which are sequences of locations to represent the patterns of trajectory. In [14], the author proposed a clustering-based approach to discover users with similar movements in a time interval. In [9] and [10], the authors use T-Pattern which is the frequent sequence of locations with similar travel time to represent movement behavior. In [25], the authors focus on inferring the transportation modes from trajectories. On the other hand, to represent individual user movement behavior, some works focus on this problem. In [12], the authors proposed a hybrid model of movement behavior for location prediction at a specific time. In [19], the authors proposed T-pattern Tree for capturing users' trajectory patterns from the trajectories of each user for next movement prediction.

One feature of the check-in data from LBSNs is spatial and temporal sparsity such that the aforementioned works do not apply to this kind of data. Some works focus on mining patterns from sparse data. In [20], the authors showed how to discover the spatio-temporal patterns from Foursquare. In [21] and [24], the authors exploited massive low-sampling-rate data to reduce the uncertainty, and used the Markov model to infer next movement. In particular, some works exploit friendships in LBSNs. In [1], the authors observed the friendship and mobility from Facebook, and proposed an algorithm for location prediction based on their observations. However, they transformed the IP to location but did not use the check-in information. In [23], the authors showed that users' mobility influenced their movement significantly, but social connection had an insignificant influence. In [6], the authors analyzed the relation between friendship and mobility, and propose two mobility model PMM and PSMM which consists of three Normal distributions with different states to model users' mobility in different time in LBSNs. However, it is hard to apply this mobility to our problem to determine the propagation probability of each edge. Therefore, we propose a distance-based mobility model DMM in this work.

B. *Diffusion Model*

Diffusion represents the spread of information when a new idea or innovation starts from nodes. A node is active when it sends a message; otherwise, it is inactive. For the viral marketing, many works focus on the diffusion model on social networks. For describing whether a node is influenced, there are two general major models, the linear threshold model (LTM) and the independent cascade model (ICM) [16]. In LTM, an activation threshold is defined; when the total influence of a node exceeds this threshold, this node becomes an active node. If a node becomes active, it will influence other connected inactive nodes. On the other hand, in ICM, when a node becomes active, it then has a probability of making its connected inactive nodes become active. This probability is independent of the influence of other connected nodes. The difference between LTM and ICM is that active nodes in ICM only have one chance to influence their connected inactive nodes, but active nodes in LTM have more than one. The above works are based on the interaction of messages. However, [18] proposed the heat diffusion model for social networks. This approach is based on time, which differs in LTM and ICM. Moreover, it is similar to LTM in terms of deciding whether the node is active or not (based on the heat of the node). However, the LBSN environment is more complex than that of traditional social networks. In [22], the authors proposed a one-wave diffusion model for LBSNs, but they did not extend the approach to traditional multi-wave consideration. Since check-in behavior can be divided into two parts, influence and check-in, using ICM to model this behavior is easier than using other models. Thus, we focus on ICM in LBSNs in this work.

C. *Influence Maximization*

The influence maximization problem is to find k users which can spread information maximization based on the diffusion model as seeds. This problem has been shown to be an NP-hard problem in general [15]. In [15] and [16], the author proposed a greedy algorithm to find the $(1 - 1/e - \epsilon)$-approximate seed set for information maximization on social networks based on LTM and ICM. In [8], the authors proposed a greedy algorithm based on the uncover degree of nodes based on ICM. Moreover, in [18], the authors proposed a greedy algorithm which is similar to [15] and [16] but based on their proposed heat diffusion model. For LBSNs, the authors of [22] designed a greedy algorithm to select users with a higher degree and more closely related to the location as seeds in their one-way diffusion model for LBSNs.

The concept of the greedy approach is to select the user which can influence the greatest number of users in each step. However, estimating the number of influenced users is hard to evaluate. A naive approach is to simulate the results by using the Monte Carlo method [15]. This greedy method is easy and effective. Nevertheless, the cost of the Monte Carlo method is huge. Thus, many works try to find the balance between being effective and efficient. In [4], the authors proposed a heuristic algorithm PMIA to estimate the influence results by using the local tree structures of each node to approximate the influence. On the other hand, CELF [17] exploits the submodular property to boost the traditional greedy approach.

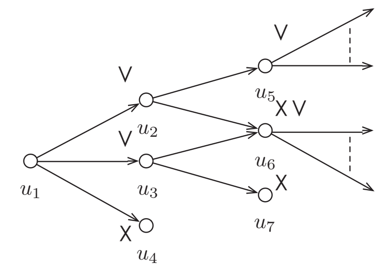


Fig. 2. An example of influence flow in LBSN using ICM

III. Problem Statement

In this section, we will give a formal definition of the location promotion problem in LBSNs. Then, we show the transformation from the location promotion problem in LBSNs to the influence maximization problem on a graph. Furthermore, we will take some naive solutions for this problem.

A. *Problem Definition*

For the location promotion in LBSNs, first, we need to define an LBSN before formally defining the problem conveniently. The formal definition of an LBSN is as follows,

Definition 1: An LBSN $\langle G, C\rangle$ consists of a social network $G = \langle V, E\rangle$, where $V = U$ which is the set of users, $E = \{(u_i, u_j)|u_j \text{ follows } u_i, u_i, u_j \in U, u_i \neq u_j\}$ and check-in set $C = \{(u, \ell, t)\}$, where (u, ℓ, t) represents a check-in where a user u checks in to location ℓ at time t, and $\ell \in L$. A location ℓ is a coordinate which consists of latitude and longitude.

Figure 1 shows an example of location promotion on an LBSN. The LBSN has five users, u_1, u_2, $\cdots$, u_5, five social connections between users, such as (u_1, u_2), and five check-ins, such as (u_1, ℓ_1, t_1). These five users check-in in three distinct locations, ℓ_1, ℓ_2, and ℓ_3. Then, the formal definition of location promotion on an LBSN is as follows,

Definition 2: Given an LBSN $\langle G, C\rangle$, a selected location ℓ and a constant k, the location promotion problem is to select a set of seeds S, $S \in U$, which has k seeds (to distinguish from other users) to maximize $\sigma(S)$, which is the number of expected users who will visit the selected location ℓ influenced by seeds.

Figure 2 shows an example of influence flow in LBSN using ICM. First, we suppose that u_1 is a seed, and then u_1's friends, u_2, u_3 and u_4, will receive a notification about check-in at the selected location from u_1. After u_1's friend u_2 receiving this notification, u_2 checks in at the selected location, but u_1's friend u_4 does not. u_2 and u_3 check in at the selected location influenced by their mutual friend u_1, then another their mutual friend u_6 will receive two notification about check-in at the selected location from u_2 and u_3. u_6 will decide whether visit the selected location two times based on his/her mobility. In this example, u_6 decides to check-in on the selected location in the second decision, and he/she will also send a notification to his/her friends which are not influenced after he/she check-ins on the selected location. The definition of the location promotion problem model is as follows,

Algorithm 1 : AF

Input: a LBSN $G = (U, E)$, k, and selected location ℓ
Output: S, which consists of k seeds

```
1:  S = ∅
2:  j = 0 # nearest location of ℓ
3:  C = ∅ # candidate set
4:  for each i from 1 to k do
5:      if C == ∅ then
6:          j = j + 1
7:          ℓ' = j-nearest location of ℓ, ell' ∈ L
8:          C = {u|(u, ℓ', t) in check-in set, ∀t}
9:      end if
10:     u = the most active user in C
11:     C = C − {u}
12:     S = S ∪ {u}
13: end for
```

Definition 3: Given an LBSN $\langle G, C \rangle$, a selected location ℓ and a constant k, the corresponding influence maximization problem on a graph is that given a graph $G = \langle V, E \rangle$ and the propagation probability of each edge $(u_i, u_j) = P_{u_i}(\ell)$, select a set of seeds S, $S \in U$, which has k seeds to maximize $\sigma(S)$, which is the number of expected active users.

B. Baseline 1: Active First

To promote a location in LBSNs, the first naive idea is to select active users who visited the selected location or near selected location. These active users have many social connections. To select these active users as seeds can help us to disseminate the information regarding the selected location since they have many social connections. Therefore, this method is to select top top-k users who have visited the selected location or near the selected location, and who have the greatest number of social connections. algorithm AF reflects this concept.

In algorithm AF, we select the nearest location of ℓ in L (line 6), and collect who visited this location from the check-in set (line 8) Then, we select the most active user in the candidate set C (line 10). The location is selected from near to far (from line 6 to line 7). For example, in Figure 1, to select one seed to promote location ℓ_3. The nearest location of ℓ_3 is itself, and the most active user who visited ℓ_3 is user u_4. Thus, the output $S = \{u_4\}$.

C. Baseline 2: Expert First

To promote a location in LBSNs, another naive idea is to select expert users who visit the selected location frequently. These expert users can influence their friends since they are experts regarding the selected location. Therefore, they have a better reputation to let their friends trust them. Thus, the method is to select the top-k users who visit the selected location or who are near selected location frequently. algorithm EF reflects this concept.

Algorithm EF is similar to algorithm AF. The only difference is in line 10, algorithm EF selects the most expert user from candidate set C as the seed (line 10). For example, in Figure 1, to select one seed to promote location ℓ_2, the nearest location of ℓ_2 is itself, and the most active user who visited ℓ_2 is user u_3. Thus, the output $S = \{u_3\}$.

Algorithm 2 : EF

Input: a LBSN $G = (U, E)$, k, and selected location ℓ
Output: S, which consists of k seeds

```
1:  S = ∅
2:  j = 0 # nearest location of ℓ
3:  C = ∅ # candidate set
4:  for each i from 1 to k do
5:      if C == ∅ then
6:          j = j + 1
7:          ℓ' = j-nearest location of ℓ, ell' ∈ L
8:          C = {u|(u, ℓ', t) in check-in set, ∀t}
9:      end if
10:     u = the most expert user in C
11:     C = C − {u}
12:     S = S ∪ {u}
13: end for
```

D. Datasets Description

In this paper, we have selected the gowalla and brightkite datasets [6] for observation and evaluation. There are 196,591 users, 950,327 social connections and 6,442,890 check-ins during February 2009 - October 2010 in the gowalla dataset. Moreover, there are 58,228 users, 214,078 social connections and 4,491,143 check-ins during April 2008 - October 2010 in the brightkite dataset.

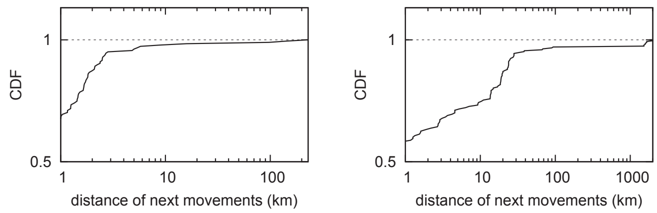
IV. Estimating Propagation Probability

To maximize the expected number of users who will visit the selected location, we have to infer whether a user will visit a specific location. Therefore, in this section, we have to capture the movement behavior of users from the check-in history of the LBSN. According to prior works [6][23], a user will visit a location due to not only social influence but also self movement behavior. Moreover, prior works show that a user's behavior is of critical importance when users visit a location on an LBSN [23]. Therefore, the user's moving behavior should be derived from their check-in data to build the mobility profile of each user. Due to the spatial and temporal sparsity of check-in data, it is hard to evaluate the probability of a user visiting a location using a small amount of movement history (only $10 - 50$ GPS records for each user on average). Thus, we exploit a two-layer model DMM instead of using massive movement history (> 100 GPS records for each user) to model user mobility.

A. Distance-based Mobility Model, DMM

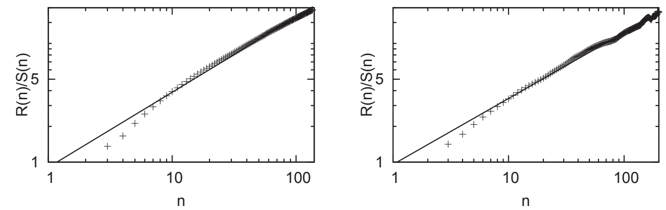
Due to the small amount of movement history of each user, it is hard to use massive movement history to estimate the probability of a user visiting a location [12][13][19]. To estimate a user's mobility, our concept of DMM is to estimate stationary distribution of observations (visited locations), then estimate the probability of the user moving from observations to the selected location. According to the discussion mentioned above, we can use the following equations to represent them. Given a location ℓ, the DMM of user u_i can be represented as Equation 1.

$$p_{u_i}(\ell) = \sum_{l} p_l^{(i)} f^{(i)}(d(l, \ell)) \tag{1}$$



(a) gowalla (b) brightkite

Fig. 3. The distribution of distance of next movements of two users in different LBSNs



(a) gowalla (b) brightkite

Fig. 4. RS test results of distance of next movements of two users in different LBSNs for estimating the Hurst exponent

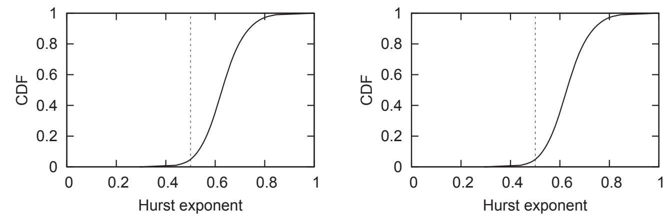
In Equation 1, $p_{u_i}(\ell)$ represents the probability density function of user u_i visiting ℓ. $p_l^{(i)}$ denotes the stationary probability of user u_i staying in location l, and $f^{(i)}(d(\ell,l))$ denotes the probability density of user u_i moving from location l to location ℓ, where $d(l,\ell)$ denotes the distance between location l and location ℓ. To estimate the $p_\ell^{(i)}$ in the LBSN of user u_i, we exploit random walk with restart to emulate a user moving in LBSNs. It is useful to capture the stationary distribution of visited locations when the data is spatially and temporally sparse [5].

To estimate $f^{(i)}(d(l,\ell))$, we need to select a suitable probability distribution for function $f^{(i)}$. However, $f^{(i)}$ represents the movement of user u_i. To select a suitable function for $f^{(i)}$, we observe the next movement of the user on the LBSN, since the distance of the next movement represents the user's mobility. We select the users which have the top-10% check-in data on the gowalla and brightkite datasets to represent normal users in LBSNs. They have 136 and 200 check-ins in the gowalla and brightkite datasets, respectively. Figure 3 shows the distributions of the next movement distance in the gowalla and brightkite datasets. Figure 3(a) and Figure 3(b) also show the long-tail property in the two LBSNs. To demonstrate our hypothesis, we selected the Hurst exponent [2] to test whether the distributions of distance of the next movement of the user is self-similar. If the distribution is self-similar, then the distribution has the long-tail property. Furthermore, we can exploit the known long-tail distribution to describe the distribution of movement distance. Figure 4 shows the results using the R/S method [2], the Hurst exponents of these two users are 0.63 and 0.56, respectively. It shows that the distance of the next movement of these two users has the self-similar property (Hurst exponents are greater than 0.5). After observing the Hurst exponent of users which have the top-20% check-ins in the LBSNs, Figure 5 shows about 95% of the top-20% of users whose distance of next movements have the self-similar property in the two LBSN datasets.

With the above observations, we utilize Pareto distribution as the function $f^{(i)}$ since it has the self-similarity property [7]. To fit the Pareto distribution of each user, suitable parameters of Pareto distribution should be selected based on their check-in history. The PDF of Pareto distribution is shown as follows:

$$p(x;\alpha,\beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$$

where α denotes the shape parameter and β denotes the location parameter. α controls the form of shape, and β controls the minimum of x. Since the x represents the distance of next movement, we fix β to 1 ($\beta \neq 0$). Then it is comparatively



(a) gowalla (b) brightkite

Fig. 5. The distribution of the Hurst exponent for each user's next movements on the LBSNs

easy to fit the distribution of shifted distance of next movement for each user since only one remaining parameter needs to be estimated.

To estimate a suitable α for each user, we exploit the maximum likelihood concept to estimate α from the check-ins of each user. The objective function is as follows:

$$\prod_i^N \mathcal{L}(\alpha; x_i, \beta=1) = \prod_i^N \frac{\alpha}{x_i{}^{\alpha+1}}$$

where $x_i = y_i + 1$, and y_i is the distance of the i-th next movement of a user. Then to find the α to maximize the objective function, the close form of α is derived in Equation 2.

$$\frac{d}{d\alpha}\prod_i^N \frac{\alpha}{x_i^{\alpha+1}} = 0, \text{ where } x_i > 0, \forall i$$

$$\prod_i^N \frac{\alpha}{x_i^{\alpha+1}} \cdot [\frac{d}{d\alpha}\sum_i^N \ln\frac{\alpha}{x_i^{\alpha+1}}] = 0$$

$$\frac{N}{\alpha} = \sum_i^N \ln x_i$$

$$\alpha = \frac{N}{\sum_i^N \ln x_i}, \text{ where } \sum_i^N \ln x_i \neq 0 \tag{2}$$

B. Propagation Probability

After deriving α for each user, $f^{(i)}(\cdot)$ in Equation 1 can be determined by α in Equation 2. Thus, $f^{(i)}(\cdot)$ can be represented as Equation 3.

$$f^{(i)}(d(\ell_1,\ell_2)) = p(d(\ell_1,\ell_2)+1;\alpha,\beta=1)$$

$$= \frac{\alpha}{(d(\ell_1, \ell_2) + 1)^{\alpha+1}} \tag{3}$$

where $d(\ell_1, \ell_2)$ denotes the distance between location ℓ_1 and ℓ_2.

To determine the propagation probability $P_\ell(u_i)$, we evaluate the probability of user moving exceeding the distance between the selected locations to each visited location and combine them using DMM in Equation 1. It describes that the user prefers to move farther than the selected location. Thus, $P_\ell(u_i)$ can be shown as Equation 4.

$$P_\ell(u_i) = \sum_l p_l^{(i)} \int_{d(l,\ell)}^{\infty} f^{(i)}(x)\mathrm{d}x$$

$$= \sum_l p_l^{(i)} \int_{d(l,\ell)+1}^{\infty} p(x; \alpha, \beta = 1)\mathrm{d}x$$

$$= \sum_l p_l^{(i)} (d(l, \ell) + 1)^{-\alpha} \tag{4}$$

V. Performance Evaluation

In this section, extensive experiments are conducted to evaluate the effectiveness of our proposed mobility model DMM and to show the simulation results of ICM. We implemented the proposed algorithms in Python.

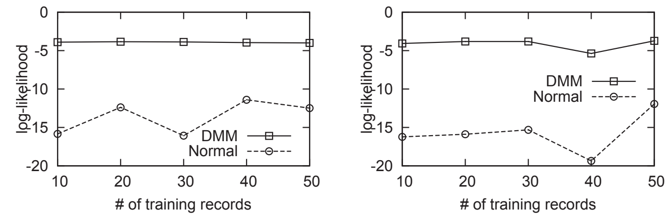
A. Comparison of Different Mobility Models

In this section, we compare the proposed mobility model DMM with Normal, a general mobility model, Normal distribution $p((x, y)) = \mathcal{N}(\mu, \Sigma)$ [11]. We split the check-in history of each user into two parts, the training part and the test part, with proportions of 80% and 20%, respectively. To evaluate the different mobility models, the metric is log-likelihood in this section. Log-likelihood reflects the quality of the mobility model. If the value of log-likelihood is higher, it shows that the quality of the mobility model is higher. To observe the performance of the two different models for spatial and temporal sparse data, we select different number of training recorded $10, 20, \cdots, 50$ in the two datasets.

Figures 6(a) and 6(b) show the results of comparison. In the gowalla dataset, DMM performs better than Normal. DMM shows similar results in different training records. It reflects that DMM can capture each user's mobility even that they have only ten check-in records. On the other hand, the performance of Normal is positively related to the number of training records. It shows that Normal needs a lot of data to capture users' mobility. Thus, Normal is not appropriate to model each user's mobility in LBSNs. Moreover, the results of brightkite dataset are similar to the results of gowalla dataset.

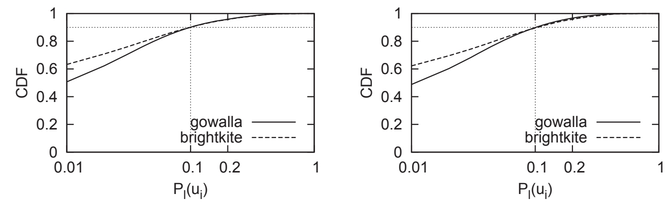
B. Propagation Probability

In this section, we observe the distribution of propagation probability of each edge derived by DMM. In general, the propagation probability in social network is set to 0.1 or $1/deg(v)$. We assume that if a propagation probability lager than 0.1, the information is easy to pass through this edge. Figure 7 shows the distribution of propagation probability at two different locations. In 7(a), the selected location is a railway station in San Francisco. About 10% users check-in at this location in the two datasets. Figure 7(a) show that about 10% users' propagation probability are larger than 0.1, the results in two datasets are similar. Furthermore, in Figure 7(b), the selected location is a park in New York city. In this location, it shows similar results to the railway station in San Francisco. The results indicate that we can generate proper propagation probability using DMM for location promotion problem.



(a) gowalla (b) brightkite

Fig. 6. The results of comparing two mobility models in different LBSNs



(a) 37.7764N 122.3943W (CA) (b) 40.7806N 73.9680W (NY)

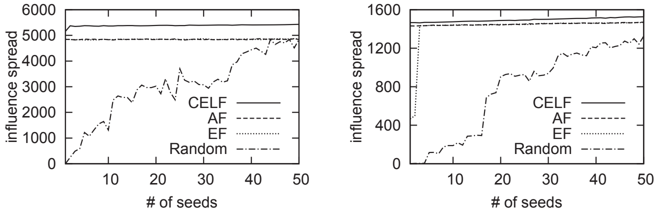
Fig. 7. The distribution of $P_\ell(u_i)$ in different LBSNs

C. Simulation Results

In this section, we show the results of the different seed selection methods. We selected a boost greedy algorithm CELF [17] for influence maximization to compare with our baseline approaches, AF, EF and Random. To evaluate the results, The Monte Carlo method is selected to evaluate the influence of a seed set $\sigma(S)$, and the number of iterations is 1000.

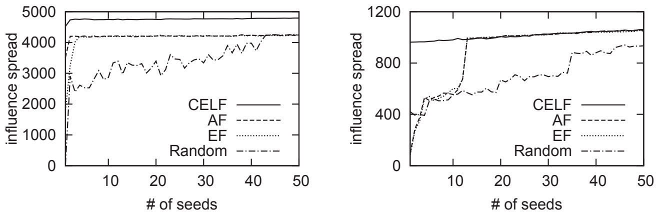
Figures 8 and 9 show the results from 50 seeds with four different selection algorithms at two locations on two datasets. All the results show that CELF gets the highest influence spread in most cases, and the AF and EF place second. Although AF and EF are not the best solution, the difference of influence between AF/EF and CELF is small. The reason is that if users visited around of the selected location, and they have higher chance to visit the selected location based on their mobility. Moreover, users having social connection in LBSN represents that they probably have similar mobility. Therefore, the seeds selected by AF/EF will influence their friends with similar mobility who probably visit the selected location.

In Figures 8(b) and 9(a), EF gets low influence spread when the number of seeds is small. It reflects that some users check-in frequently near the selected location , but they have low degree of outcome social connections. It limits the influence spread when the degree of outcome social connections is low. In 9(b), the gap between AF/EF and CELF is large when the number of seeds is less than 12. The reason is that these seeds are tourists, they frequently visit and check-in near the



(a) gowalla (b) brightkite

Fig. 8. The influence spread of different algorithms promoting location 37.776430N 122.394318W (CA) in different LBSNs



(a) gowalla (b) brightkite

Fig. 9. The influence spread of different algorithms promoting location 40.780606N 73.968088W (NY) in different LBSNs

selected location (Central Park, NYC), but their mobility are not around this location. Thus, their friends would not like to visit this location since their mobility are not near the selected location. It shows that some native users may not check-in at some scenic spots in their city since they are accustomed. It causes that the frequent of check-ins by tourists is higher than the frequent of check-ins by townspeople at some scenic spots. In all cases, CELF gets the largest influence spread since CELF selects more influential users as seeds. Our results also demonstrate that the location promotion problem can be modeled as an influence maximization problem and solved by the methods for influence maximization on ICM.

VI. Conclusion

In this paper, we modeled the location promotion problem in LBSNs as an influence maximization problem on ICM in a graph. To determine the propagation probability of each edge, we proposed a distance-based mobility model DMM for each user on an LBSN. DMM utilizes random walk with restart and Pareto distribution to deal with the spatial and temporal sparsity of the check-in data of LBSNs. The experimental results show that DMM can capture each user's mobility in LBSNs effectively, and DMM can generate proper propagation probability to describe the influence flow in LBSNs for location promotion problem.

Acknowledgment

This research was supported in part by the Ministry of Science and Technology of Taiwan (MOST 102-2221-E-009-171-MY3, MOST 102-2219-E-001-002, and MOST 100-2218-E-009-016-MY3), Academia Sinica (AS-102-TP-A06), Taiwan MoE ATU Program, and Microsoft.

References

[1] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity," in *WWW*, April 2010.

[2] J. Barunik and L. Kristoufek, "On hurst exponent estimation under heavy-tailed distributions," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 18, pp. 3844–3855, 2010.

[3] H. Cao, N. Mamoulis, and D. W. Cheung, "Mining frequent spatio-temporal sequential patterns," in *IEEE ICDM*, November 2005.

[4] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *ACM SIGKDD KDD*, July 2010.

[5] M.-F. Chiang, W.-Y. Zhu, W.-C. Peng, and P. S. Yu, "Distant-time location prediction in low-sampling-rate trajectories," in *IEEE MDM*, June 2013.

[6] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *ACM SIGKDD KDD*, August 2011.

[7] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, December 1997.

[8] P. A. Estevez, P. Vera, and K. Saito, "Selecting the most influential nodes in social networks," in *International Joint Conference on Neural Networks*, August 2007.

[9] F. Giannotti, M. Nanni, and D. Pedreschi, "Efficient mining of temporally annotated sequences," in *SIAM SDM*, April 2006.

[10] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *ACM SIGKDD KDD*, August 2007.

[11] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, June 2008.

[12] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *IEEE ICDE*, April 2008.

[13] H. Jeung, M. L. Yiu, X. Zhou, and C. S. Jensen, "Path prediction and predictive range querying in road network databases," *The VLDB Journal*, vol. 19, no. 4, pp. 585–602, August 2010.

[14] P. Kalnis, N. Mamoulis, and S. Bakiras, "On discovering moving clusters in spatio-temporal data," in *SSTD*, August 2005.

[15] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD KDD*, August 2003.

[16] ——, "Influential nodes in a diffusion model for social networks," in *International Colloquium on Automata, Languages and Programming*, July 2005.

[17] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *ACM SIGKDD KDD*, August 2007.

[18] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," in *ACM CIKM*, November 2008.

[19] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *ACM SIGKDD KDD*, June 2009.

[20] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," in *International AAAI Conference on Weblogs and Social Media*, July 2011.

[21] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *ACM SIGKDD KDD*, August 2012.

[22] H.-H. Wu and M.-Y. Yeh, "Influential nodes in one-wave diffusion model for location-based social networks," in *PAKDD*, April 2013.

[23] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *ACM SIGIR*, July 2011.

[24] K. Zheng, Y. Zheng, X. Xie, and X. Zhou, "Reducing uncertainty of low-sampling-rate trajectories," in *IEEE ICDE*, April 2012.

[25] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *ACM UbiComp*, September 2008.