Exploiting Viral Marketing for Location Promotion in Location-Based Social Networks

WEN-YUAN ZHU, Industrial Technology Research Institute
WEN-CHIH PENG, National Chiao Tung University
LING-JYH CHEN, Academia Sinica
KAI ZHENG, Soochow University
XIAOFANG ZHOU, The University of Queensland; Soochow University

25

With the explosion of smartphones and social network services, location-based social networks (LBSNs) are increasingly seen as tools for businesses (e.g., restaurants and hotels) to promote their products and services. In this article, we investigate the key techniques that can help businesses promote their locations by advertising wisely through the underlying LBSNs. In order to maximize the benefit of location promotion, we formalize it as an influence maximization problem in an LBSN, i.e., given a target location and an LBSN, a set of k users (called seeds) should be advertised initially such that they can successfully propagate and attract many other users to visit the target location. Existing studies have proposed different ways to calculate the information propagation probability, that is, how likely it is that a user may influence another, in the setting of a static social network. However, it is more challenging to derive the propagation probability in an LBSN since it is heavily affected by the target location and the user mobility, both of which are dynamic and query dependent. This article proposes two user mobility models, namely the Gaussian-based and distance-based mobility models, to capture the check-in behavior of individual LBSN users, based on which location-aware propagation probabilities can be derived. Extensive experiments based on two real LBSN datasets have demonstrated the superior effectiveness of our proposals compared with existing static models of propagation probabilities to truly reflect the information propagation in LBSNs.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Data mining, spatial databases and GIS*

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Propagation probability, influence maximization, check-in behavior, location-based social network

ACM Reference Format:
Wen-Yuan Zhu, Wen-Chih Peng, Ling-Jyh Chen, Kai Zheng, and Xiaofang Zhou. 2016. Exploiting viral marketing for location promotion in location-based social networks. ACM Trans. Knowl. Discov. Data 11, 2, Article 25 (November 2016), 28 pages.
DOI: http://dx.doi.org/10.1145/3001938

This research was supported in part by the Ministry of Science and Technology, Taiwan (104–2221–E–009–138–MY2), Academia Sinica (AS–105–TP–A07), Taiwan MoE ATU Program, National Natural Science Foundation of China (Grant 61472263) and Australian Research Council (DP120102829 and LP130100164).
Authors' addresses: W.-Y. Zhu, Industrial Technology Research Institute, Hsinchu 31040, Taiwan; email: wyzhu@itri.org.tw; W.-C. Peng, Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan; email: wcpeng@cs.nctu.edu.tw; L.-J. Chen, Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan; email: cclljj@iis.sinica.edu.tw; K. Zheng, School of Computer Science and Technology, Soochow University, Suzhou 215006, China; email: zhengkai@suda.edu.cn; X. Zhou, School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia QLD 4072, Australia, and School of Computer Science and Technology, Soochow University, Suzhou 215006, China; email: zxf@itee.uq.edu.au.

DOI: http://dx.doi.org/10.1145/3001938

1. INTRODUCTION

Due to the success of viral marketing, more and more advertisements are appearing in social networks. The key to the success of viral marketing is the influence among social connections. Users are more likely to accept advertisements from their friends in social networks than from media directly. By observing this phenomenon, prior works have elaborated on the influence maximization problem in social networks. In general, a social network is modeled as a graph $G = \langle U, E\rangle$, where U is the set of users, E refers to the social connections among users, and the weight of edges infers the influence degree between users. Given a graph, the influence maximization problem is to select a set of users as seeds with the purpose of maximizing the expected number of active users (i.e., influence spreads) in a social network [Kempe et al. 2003].

With the popularity of smart phones and location-based social networks (LBSNs), users are able to check in at some locations and share their check-in records with their friends. If their friends are interested in these locations, they may visit them. Recently, many points of interest (POIs) (e.g., restaurants and stores) have explored check-in sharing to attract users to stay or visit since the check-in action represents the visiting action as well as the spread of information. Via the check-in records of users, more users (e.g., friends of check-in users) may be interested in the POI and then visit it. If we aim to promote a location using viral marketing in an LBSN, the location promotion problem is formulated, where given one target location and the number of seeds, the purpose is to maximize the expected number of users who will check in at the target location. Intuitively, the location promotion problem can be modeled as an influence maximization problem in LBSNs. Explicitly, given a set of nodes and edges with the propagation probability, a target location and the number of seeds, a set of seed users is derived to maximize the expected number of active users, where the active users represent the users who will check in at the target location. Notice that one challenging issue associated with the location promotion problem is to determine the propagation probabilities of edges in an LBSN.

In this article, we focus on deriving the propagation probabilities based on users' check-in records in an LBSN. By investigating the information of LBSNs (i.e., check-in locations of users and social connections of users), we first utilize some existing static approaches [Kempe et al. 2003, 2005; Goyal et al. 2010; Chen et al. 2009, 2010; Bouros et al. 2014] for the propagation probabilities. However, successful information propagation in LBSNs means that users should follow their friends to check in at the target location specified. It means that the check-in action triggers the information propagated to their friends. Thus, to achieve this, we claim that the individual check-in behavior should be considered. Moreover, if the target location is changed, the propagation probability should be updated. Therefore, to infer whether a user checks in at the target location or not, we should take into account the individual check-in behavior.

Based on the above discussions, one should consider the individual check-in behavior for the location promotion problem. To the best of our knowledge, there are some mobility models to describe individual check-in behavior in LBSNs [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014]. The mobility models proposed [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014] are all based on the bivariate Gaussian distribution. However, it is hard to decide the area around the target location to evaluate the probability of a user check-in at the target location from Gaussian-based mobility models (GMMs), which are two-dimensional density functions. Moreover, the GMMs only describe the coordinates of check-in records, but do not consider their order. We argue that different orders of check-in records represent different check-in behavior. On the other hand, the models of the above prior works only utilize the coordination information of locations. In most LBSNs, the location contains not only the coordination information, but also the category information. The category information indicates the property and functionality of the corresponding location. Each user also has their categorical preference of locations [Ye et al. 2013]. Thus, the category information should be considered to describe individual check-in behavior more precisely.

To deal with the above issues, we propose distance-based mobility models (DMMs) to represent individual check-in behavior. The idea is that a user is likely to visit the target location if the user has appeared in its nearby area of the target location. For the distance-based approach, we exploit the stationary distribution of visited locations and the distance between consecutive check-in records to model individual check-in behavior, where the stationary distribution of the visited locations and distances represent the individual locality and mobility preferences, respectively. In practice, the distance-based approach can easily evaluate the probability of user check-in at the target location. Thus, our proposed DMMs are suitable, not only for describing individual check-in behavior, but also for determining the propagation probability of edges in an LBSN.

In summary, our major contributions are outlined as follows.

—We formulate the location promotion problem as an influence maximization problem in an LBSN.
—We extend the independent cascade model (ICM) to a location-aware independent cascade model (LICM) to describe the information propagation in an LBSN.
—By investigating the check-in records of users and the social connections of users, we organize and utilize some existing static approaches for the propagation probability in LBSNs.
—We take the target location and users' check-in records into consideration and propose two types of mobility model, GMMs and DMMs, to capture the individual check-in behavior as well as to evaluate location-aware propagation probability in LBSNs.
—To precisely capture the individual check-in behavior in LBSNs, we extend our Gaussian-based approach and distance-based approach by considering not only the spatial-temporal information of check-in records, but also the categorical information of locations.
—We have conducted comprehensive experiments on two real datasets, and the experimental results show that the proposed models are suitable for effective influence maximization in LBSNs.

The remainder of this article is organized as follows: Section 2 discusses related works. Section 3 outlines the preliminaries. Sections 4 and 5 present the static and location-aware approaches for the propagation probabilities in LBSNs, respectively. The experimental results are shown in Section 6. Section 7 concludes this article.

2. RELATED WORKS

In this article, our work is related to the influence maximization in social networks. Thus, we will present some existing works of propagation models and influence maximization in social networks. Since the main theme of this article is to derive the propagation probabilities, we will describe how to set the propagation probabilities in existing works. Note that as our work is to explore mobility models from check-in records to set the propagation probabilities, we will present some research works of modeling user mobility from social media. Moreover, since POI recommendation is closely related to location promotion, we will also discuss some POI recommendation works and the factors they considered in these works.

2.1. Propagation Models for Social Networks

The influence maximization is based on a given propagation model. There are two general propagation models for social networks, the ICM and the linear threshold model (LTM) [Kempe et al. 2003, 2005]. In ICM, each directed edge (u, v) has a corresponding propagation probability $pp(u, v)$. If u is active in time slot t, then v becomes active with the probability $pp(u, v)$ in time slot $t + 1$. Initially, the seeds are set active in time slot 0. Then, the propagation process stops when no user becomes active in a time slot.

To determine the propagation probability in a social network, some existing approaches utilize the number of in-degrees [Kempe et al. 2003], a fixed value [Chen et al. 2009, 2010] and a uniformly random value [Chen et al. 2009, 2010]. However, these approaches are some baseline methods. In social networks, users' actions do reflect real influences. As such, some works focus on learning the propagation probability from action history. In Saito et al. [2008], the authors exploited the EM algorithm to learn the probability in ICM. In Goyal et al. [2010], the authors utilized the Bernoulli distribution to model the influence on each social connection. The success probability parameter of the Bernoulli distribution is learned from data for every social connection, and each success probability parameter is the propagation probability of the corresponding edge.

In Tang et al. [2009], Weng et al. [2010], Liu et al. [2010], and Lin et al. [2011], the authors noticed that different topics affect the propagation since users have different topic preferences. Thus, in Barbieri et al. [2012], the authors proposed a topic-aware independent cascade model (TICM) and topic-aware linear threshold model (TLTM) based on ICM and LTM, respectively. TICM and TLTM describe the information propagation of the information with different topics. Especially, in TICM, the propagation probability $pp(u, v|\gamma)$ corresponds to the target information γ, where γ can be represented by a topic distribution $\gamma = (\gamma^1, \gamma^2, \ldots, \gamma^Z)$ if there are Z topics. Then, the definition of $pp(u, v|\gamma)$ is $pp(u, v|\gamma) = \sum_i \gamma^i pp^i_{u,v}$, where $pp^i_{u,v}$ denotes the propagation probability from u to v if the information totally belongs to the ith topic. In Barbieri et al. [2012], the authors also provided an EM-based approach to detect the topic parameters of each edge.

In LBSNs, this phenomenon is more observable. Users have their locality such that they are likely to move to a near location. Thus, different target locations have different propagation probabilities. However, the set of locations is infinite and uncountable. It is hard to distinguish locations into finite topics or areas. Therefore, the topic-aware influence models are ineligible to represent the information propagation in LBSNs. In Zhou et al. [2015], the authors proposed the TP model based on ICM to describe the online-to-offline process in LBSNs. The TP model consists of two phases, online and offline. The online phase is similar to ICM. Each edge (u, v) has a corresponding online probability, which is the probability that u influences v. The online probability is the same as that of the traditional ICM. When a user becomes active in the online stage, the online phase is switched to the offline phase. In the offline phase, this user will become offline active with an offline probability corresponding to the user and a given location, where offline probability represents the probability of the user purchasing the product at the given location. If a user becomes offline active, this user will attempt to activate his/her friends as online active with a corresponding online probability in the next time slot. However, it is hard to connect the online and offline business to the information propagation in an LBSN. Moreover, the authors do not provide a solution to find the online probability in the TP model. Therefore, we think that the TP model is not eligible to represent the information propagation in LBSNs. Due to the above discussions, in this work, we try to model the information propagation in an LBSN based on ICM and derive the propagation probability corresponding to a given target location.

2.2. Influence Maximization in Social Networks

Given a propagation model, the influence maximization problem is to find a seed set S with k users that can maximize the influence spreads $\sigma(S)$ in a social network, and this problem is NP-hard [Kempe et al. 2003]. Since σ based on ICM is monotone and submodular, the greedy approach for seed selection archives $(1 - \frac{1}{e})$-approximation [Kempe et al. 2003, 2005]. To make the greedy approach more efficient, in Leskovec et al. [2007], the authors proposed the CELF algorithm that exploits the submodular property to significantly boost the traditional greedy approach. Moreover, the authors of Goyal et al. [2011] proposed an enhanced algorithm of CELF, called CELF++, to improve the efficiency of the greedy approach. On the other hand, another computation issue is to evaluate $\sigma(S)$. A traditional way is to utilize the Monte Carlo approach [Kempe et al. 2003, 2005]. However, the computation cost of this approach is expensive since it has to run about 10,000 times for evaluation [Kempe et al. 2003]. In Chen et al. [2010], the authors indicated that evaluating $\sigma(S)$ is #P-hard. Thus, many works [Chen et al. 2009, 2010; Jung et al. 2012; Cohen et al. 2014] have proposed approximation approaches to estimate the influence spreads $\sigma(S)$ and the marginal gain[1] $\sigma(S \cup \{u\}) - \sigma(S)$. Especially, Goyal et al. [2012] proposed an approach to select seeds from activity records. For influence maximization based on topic-aware propagation models, Aslay et al. [2014] and Chen et al. [2015] tried to improve the computation efficiency of seed selection when the parameters are derived in TICM.

For LBSNs, the authors of Wu and Yeh [2013] designed a greedy algorithm to select users with a higher degree and more closely related to the location as seeds in their one-wave diffusion model for LBSNs. However, the one-wave diffusion model is not close to real LBSNs. In Li et al. [2014], the authors formulated a location-aware influence maximization problem in LBSNs. In their network model, each user has one representative location. Their problem is given a query region, to derive k seeds with the purpose of maximizing influence spreads in the query region. In reality, locations in social media are referred to as POIs that could be restaurants, hotels, or theme parks. From the perspective of POIs, each would like to attract users to visit, and via the check-in records of users, more users (e.g., friends of check-in users) would be interested in and then visit this POI. Therefore, this problem is different from the location promotion problem addressed in this article.

2.3. Capturing Individual Check-In Behavior

The check-in records can be seen as trajectories from each user's perspective since a trajectory is a sequence of locations ordered by visiting time. Many works have aimed to discover spatio-temporal patterns from trajectories [Kalnis et al. 2005; Giannotti et al. 2006, 2007; Zheng et al. 2014]. To capture an individual check-in behavior, we have to deal with the spatial and temporal sparsity. In Noulas et al. [2011], the authors discovered spatio-temporal patterns from all users' check-in records in Foursquare. In Ye et al. [2011], the authors showed that the major reason for check in is based on their mobility, and the chance of social influence is very low. In Cho et al. [2011], the authors had similar perspectives on weak social influence in LBSNs. The authors proposed PSMM, which divides each user's check-ins into three types: work, home, and social. The check-in records of each type are modeled by a bivariate Gaussian distribution, except the social type. Moreover, to distinguish the home and work states of check-in records, they also select the Gaussian distribution to classify the time of the check-in records. In Gao et al. [2013b], there are two factors considered for users' check-in

[1]The marginal gain is used in the greedy approach to select a new seed which can increase the most influence spreads.

behavior, personal preference, and social influence. Personal preference is modeled by a bivariate Gaussian distribution, and the social influence is from the similarity between the user's check-in records and their friends' check-in records. In Lichman and Smyth [2014], the authors exploit kernel density estimation (KDE) to model the individual user's check-in records. They provide how to set the parameters of KDE for each user from their check-in records. The KDE approach is Gaussian based since it is a mixture model consisting of bivariate Gaussian distributions. In our work, it is hard for the models to evaluate the probability of a location since we have to decide on an area to compute the probability (the probability of a point in a two-dimensional distribution is 0). Therefore, we propose a distance-based approach to model individual check-in behavior. Similarly, the authors in Zhou et al. [2015] also proposed a distance-based approach to evaluate the probability that a user goes to the given location to shop based on check-in records. This approach utilizes the truncated power law [González et al. 2008] to model the distance between consecutive check-in records using all check-in records. However, this approach does not take into account the user preference since different users have different check-in behavior.

2.4. Point-of-Interest Recommendation in LBSNs

The POI recommendation is a type of recommendation system that focuses on recommending users some locations they may be interested in and would like to visit. In LBSNs, the recommendation systems should know users' locations of interest based on check-in records. Due to the popularity of the collaborative filtering (CF) technique in recommendation systems, some CF-based POI recommendation systems are proposed [Ye et al. 2010; Ference et al. 2013]. The concept of the CF-based approach is to recommend the locations based on users with similar visited locations. Since the LBSN links the online and offline activities, the POI recommendation is more complex than the normal product recommendation. Thus, only considering the concept of CF is not enough in POI recommendation. Since users would like to visit close locations, several works [Ye et al. 2011; Cheng et al. 2012; Lian et al. 2014; Hu and Ester 2014] consider the geographical feature in their POI recommendation systems. Moreover, some works also consider the popularity feature [Ye et al. 2011; Liu et al. 2013; Yin et al. 2013]. Some users will be attracted to visiting popular locations, such as tourist attractions. To take into account the temporal behavior of users and locations, some works [Cheng et al. 2013; Gao et al. 2013a; Yuan et al. 2013; Hu and Ester 2014] adopt the temporal feature to describe the check-in preference. In LBSNs, users are able to receive their friends' check-in activities. Thus, users may be attracted to visit the locations their friends have visited. Some works [Ye et al. 2011; Noulas et al. 2012; Ye et al. 2012] consider the social influence in their works. However, Ye et al. [2011] indicates that social influence does play an important role in user check-in behavior. In LBSNs, most locations have one or more corresponding categories. Moreover, some check-in information contains text information, so we can derive some semantic information about the locations. Thus, some content-based approaches [Hu and Ester 2013; Liu and Xiong 2013; Yin et al. 2013; Gao et al. 2015; Wang et al. 2015] are proposed. POI recommendation and location promotion both infer the interesting locations for users, but there are some differences between POI recommendation and location promotion. For POI recommendation, only visited locations are adopted to recommend to users. However, for location promotion, the target locations are usually new, so these locations should be promoted. Thus, the temporal behavior of locations is hard to gather in location promotion. In addition, for location promotion in this article, we do not care when each user checks in at a target location since we focus on whether the information propagates or not. According to the above discussion, for location promotion in this article, we focus on the geographical feature as well as the semantic feature to infer the probability of check in as well as the propagation probability, and the temporal and social features are not our concern in this article.

3. PRELIMINARIES

In this section, we first give a formal definition of the location promotion problem and the main problem we focus on in this article. Then, we show the framework of location promotion in an LBSN and justify our proposed framework.

3.1. Influence Maximization in LBSNs

In this article, the main platform of location promotion is the LBSN. The formal definition of an LBSN is as follows.

Definition 3.1 (*LBSN*). An LBSN $\langle G, C\rangle$ consists of a social network $G = \langle U, E\rangle$, where U is the set of users, $E = \{(u_i, u_j)|$ a social connection from u_i to u_j, $u_i, u_j \in U, u_i \neq u_j\}$ and the set of check-in records $C = \{(u, \ell, t)\}$, where (u, ℓ, t) represents a check-in record where a user u checks in at location ℓ at time t. A location ℓ is a coordinate that consists of latitude and longitude.

Then, the formal definition of the location promotion problem is as follows.

Definition 3.2 (*Location Promotion Problem*). Given an LBSN $\langle G, C\rangle$, a target 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 the expected number of users who will visit the target location ℓ, denoted as $\sigma(S|\ell)$.

The location promotion problem can be formulated as an influence maximization problem in an LBSN. To evaluate the influence spread $\sigma(S|\ell)$, we extend ICM to describe the information propagation in an LBSN. The LICM based on ICM is devised to describe the information propagation in an LBSN. The definition of LICM is as follows.

Definition 3.3 (*LICM*). Given a directed social graph $G = \langle U, E\rangle$ and a seed set S and a target location ℓ, for all $s \in S$ will spread the check-in information of ℓ (becoming active) in time slot 0. If a user u is active in time slot t, then v check-in at ℓ (becoming active) with the probability $pp(u, v|\ell)$ in time slot $t+1$, where $\exists(u, v) \in E$. The activation attempts are independent, and the activation attempt is only effective in one time slot.

According to the discussion in Section 1, we think that the propagation probability in an LBSN is related to the target location. To describe the information propagation in an LBSN, the definition of the propagation probability in an LBSN is as follows.

Definition 3.4 (*Propagation Probability in LBSNs*). The propagation probability of edge (u, v) with respect to the target location ℓ denoted as $pp(u, v|\ell)$. $pp(u, v|\ell)$ is the probability of the check-in information of ℓ propagating from u to v.

Therefore, $pp(u, v|\ell)$ is also the probability of v checking in at ℓ if u checked in at ℓ in the last time slot.

3.2. Framework

When the target location ℓ is given, the behavior of LICM is the same as the behavior of ICM. Lemma 3.5 is shown as follows.

LEMMA 3.5. *Given a target location ℓ, the properties of LICM are the same as the properties of ICM.*

PROOF. When ℓ is given, $pp(u, v|\ell)$ is a constant, and then $pp(u, v) = pp(u, v|\ell)$ in ICM. Thus, LICM and ICM have the same properties when ℓ is given in LICM. □

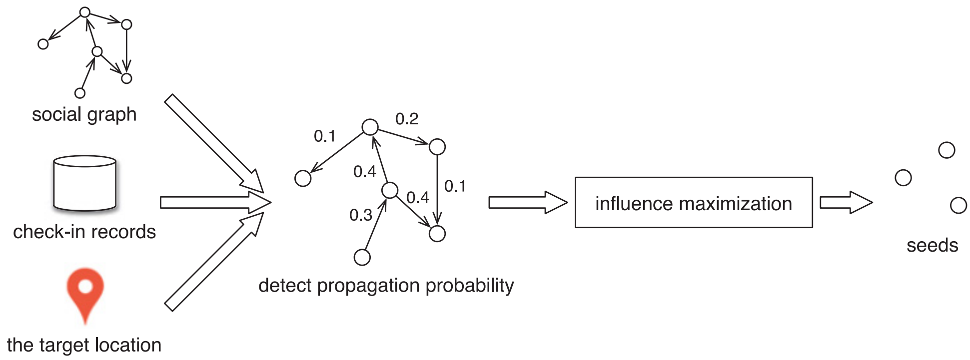


Fig. 1. The framework for location promotion in an LBSN.

Based on Lemma 3.5, the greedy approach for seed selection in LICM is $(1 - \frac{1}{e})$-approximation, and it is #P-hard to evaluate the influence spreads based on LICM. Thus, to estimate the influence spread and the marginal gain based on LICM, the approximation approaches for ICM can directly apply to LICM, such as in Chen et al. [2009, 2010], Jung et al. [2012], and Cohen et al. [2014].

Figure 1 shows the framework of location promotion in an LBSN. In the first step, given a target location, we determine the propagation probability of edges in the social graph based on check-in records. In this step, the input of location promotion in an LBSN is transformed to the input of influence maximization on a graph. Based on the social graph and propagation probability of edges from the first step, one could derive the seed set via existing solutions of influence maximization, such as the greedy approaches for seed selection [Leskovec et al. 2007; Goyal et al. 2011] and the approximations for the influence spreads [Chen et al. 2009, 2010; Jung et al. 2012; Cohen et al. 2014]. One challenging issue is how to set the propagation probability on social connections to truly reflect the propagation of information in an LBSN. The following sections will present how to derive the propagation probability of each edge in an LBSN.

4. STATIC APPROACHES

In this section, we show the traditional approaches for the propagation probability on edges in a social network. By referring to prior works [Kempe et al. 2003; Saito et al. 2008; Chen et al. 2009, 2010; Goyal et al. 2010, 2012], we organize and show the approaches to derive the propagation probability[2] $pp(u, v)$ of edge (u, v) as follows.

Uniform: All edges are assigned to the same probability 0.01 [Goyal et al. 2012].

Trivalency: All edges are assigned to the probability selected from $\{0.1, 0.01, 0.001\}$ uniformly [Chen et al. 2010; Goyal et al. 2012].

In-degree: The propagation probability from u to v is the in-degree of v [Kempe et al. 2003; Chen et al. 2009, 2010; Goyal et al. 2012]

$$pp(u, v) = \frac{1}{i\text{-}deg_v},$$

where $i\text{-}deg_v$ denotes the in-degree of v. The probability is higher if the in-degree of v is lower.

Jaccard index of friends: The Jaccard index is used to measure the similarity between two sets. Here, we select the Jaccard index to measure the similarity between two users'

[2]In this section, $pp(u, v|\ell)$ abbreviates to $pp(u, v)$ since the prior static approaches are irrelevant to ℓ.

friendships as the propagation probability if the edge (u, v) exists [Goyal et al. 2010]

$$pp(u, v) = \frac{|F_u \cap F_v|}{|F_u \cup F_v|},$$

where $F(u)$ denotes the set of u's friends. The probability is higher if the proportion of common friends is higher.

Jaccard index of locations: For LBSNs, we selected the Jaccard index of locations from two users' check-ins as the propagation probability if the edge (u, v) exists [Bouros et al. 2014]

$$pp(u, v) = \frac{|L_u \cap L_v|}{|L_u \cup L_v|},$$

where L_u denotes the set of u's visited locations. The propagation probability is higher if most visited locations of u and v are the same.

Cosine of locations: To consider the check-in times of locations, we convert the user's check-in records into a vector, in which the dimensions are locations and the values of each dimension are the check-in times of this location. Thus, we can measure the similarity via cosine from two users' check-in records

$$pp(u, v) = \frac{\sum_{\ell \in L_u \cap L_v} \#_{u,\ell} \cdot \#_{v,\ell}}{\sqrt{\sum_{\ell \in L_u} (\#_{u,\ell})^2} \sqrt{\sum_{\ell \in L_v} (\#_{v,\ell})^2}},$$

where $\#_{u,\ell}$ denotes the number of times u visited ℓ. The propagation probability is higher if not only most visited locations of u and v are the same, but also the check-in times of the visited locations are similar.

EM: To estimate the propagation probability based on ICM, Saito et al. [2008] utilized the EM algorithm to learn the best propagation probability on edges based on ICM from check-in records.

Bernoulli: Given two check-in records of two users connected by a social connection, we can calculate the times of successful propagation. For each check-in (u, ℓ, t), we can know whether u's friends have checked in at the same location after time t. If yes, then the propagation succeeds; otherwise, it fails. The probability of successful propagation is from the Bernoulli distribution. To find the success probability in the Bernoulli distribution from the records, the result from the maximum likelihood estimator is as follows [Goyal et al. 2010]:

$$pp(u, v) = \frac{\#_{u,v}}{\#_u},$$

where $\#_{u,v}$ and $\#_u$ denote the times of successful propagation from u to v and the number of attempts from u, respectively. The probability is higher if the time of successful propagation from u to v is higher.

5. LOCATION-AWARE APPROACHES

In LBSNs, the information propagation is different from in traditional social networks since users have to perform check-in at the target location to spread the information.[3] Figure 2 shows an example of information propagation in an LBSN. Given the target location ℓ, u_1 is active such that u_1's friend u_2 will receive the information of the

[3]In this article, we exploit the feature of check-in action for location promotion, where the check-in action means that the location information has been propagated and the user has been at the location. However, the retweeting or liking action does not contribute directly to location promotion in LBSNs. Thus, in this article, we only take the check-in action into consideration.

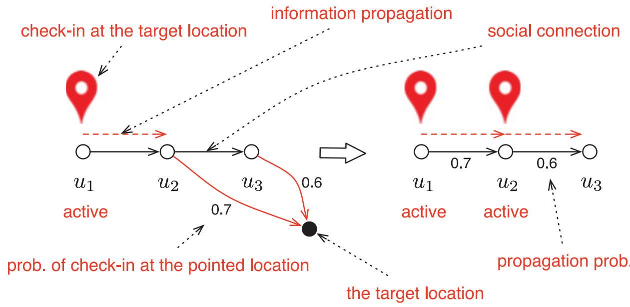


Fig. 2. An example of information propagation in an LBSN.

target location from u_1's check-in sharing. Based on Definition 3.4, $pp(u_1, u_2|\ell)$ is the probability of u_2 checking in at ℓ after u_2 receives the information of ℓ from u_1, denoted as $pr(\ell|u_2, \text{from} = u_1)$. If u_2 does check in, the information of ℓ will propagate to u_2's friend, u_3. The situation can be represented by the following formula:

$$pp(u_1, u_2|\ell) = pr(\ell|u_2, \text{from} = u_1). \tag{1}$$

Prior works [Cho et al. 2011; Ye et al. 2011] indicated that social influence is insignificant in LBSNs. Only 10%–20% of check-in records are influenced by friends. Therefore, given a target location ℓ, we assume that u is a trigger to let v know ℓ if $(u, v) \in E$ and u is active. If v is interested in ℓ, v will check in at ℓ based on their locality preference. Hence, Equation (1) can be simplified to the following formula:

$$pp(u_1, u_2|\ell) = pr(\ell|u_2, \text{from} = u_1) \approx pr(\ell|u_2), \tag{2}$$

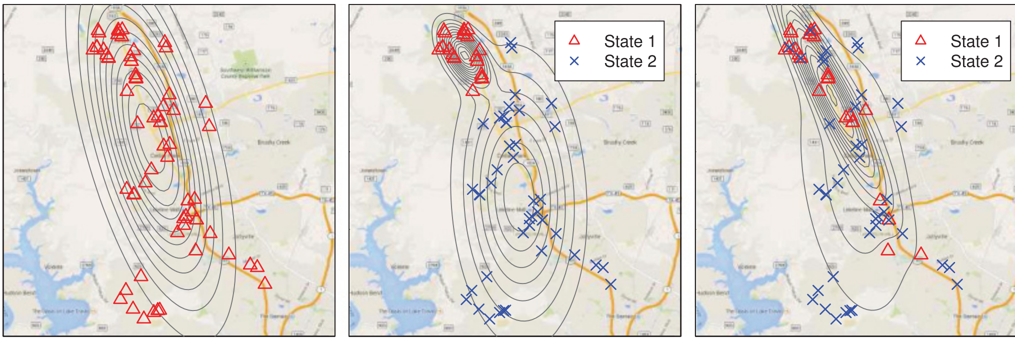
where $pr(\ell|u_2)$ denotes the probability of u_2 checking in at ℓ. Based on Equation (2), given v, for all u, $pp(u, v|\ell)$ has the same value if $(u, v) \in E$.

Although the propagation probabilities of a destination user should be different from different source users, we consider that the propagation probabilities of a destination user are the same from different source users in our approach since the influence factor in LBSNs is insignificant [Cho et al. 2011; Ye et al. 2011]. To compute the influence of the source users, the source's check-in records should be considered. However, it is unwise to compute an insignificance factor for each edge. Thus, the influence from source users in propagation probability is omitted in our approach.

In Figure 2, $pr(\ell|u_2)$ and $pr(\ell|u_3)$ are 0.7 and 0.6, respectively. Thus, $pp(u_1, u_2|\ell)$ and $pp(u_2, u_3|\ell)$ are 0.7 and 0.6, respectively. Clearly, if the target location is changed, the probabilities will also be changed. To evaluate $pr(\ell|u)$, we explore the mobility model[4] in an LBSN to capture individual check-in behavior as well as to infer the propagation probability in LBSNs. There are two kinds of approaches proposed: one is the Gaussian-based approach and the other is the distance-based approach.

In the prior works [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014; Zhu et al. 2015], the location ℓ only includes the coordinate information (i.e., latitude and longitude). In most LBSNs, each location has a profile page. The profile page contains other rich information, such as category information, photos, text descriptions, and user comments. Usually, the location has a corresponding category information, such as restaurant, shops, and transpiration. The category information represents the function and property of the location [Ying et al. 2010; Bao et al. 2012; Ye et al. 2013]. Users also have their category preference in LBSNs. Thus, we will consider not only

[4]The mobility model we focus on in this article is a PDF $p(\ell|u)$.



(a) GMM-Basic (b) GMM-Spatial (c) GMM-Temporal

Fig. 3. Examples of Gaussian-based mobility models.

the coordinate information, but also the category information of location to estimate $pr(\ell|u_2)$ in Equation (2).

5.1. Gaussian-Based Approaches

Prior works [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014] utilized the GMMs to model individual check-in behavior in LBSNs. However, in the prior works [Cho et al. 2011; Gao et al. 2013b], their mobility models focused on $p(\ell|u,t)$, which denotes that the PDF of u checks in at ℓ at time t. In our work, we do not care about the probability of a location at a specific time since we just care about whether each user checks in at the target location or not.[5] Thus, we show three major types of GMMs based on existing works [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014].

GMM-Basic: This is the basic form of GMM. Each user is modeled by one bivariate Gaussian distribution. Given the check-in records $\{(u, \ell = (x, y), t)\}$ of user u, u's mobility model is as follows:

$$p(\ell|u) = \mathcal{N}(\mu_u, \Sigma_u) = \frac{1}{2\pi\sqrt{|\Sigma_u|}} \exp\left(-\frac{1}{2}(\ell - \mu_u)^T \Sigma_u^{-1}(\ell - \mu_u)\right), \tag{3}$$

and the parameters of $p(\ell|u)$ can be estimated by the following equations: $n = |\{(u, \ell = (x, y), t)\}|$, $\hat{\mu_u} = \frac{1}{n}\sum[x, y]$ and $\hat{\Sigma_u} = \frac{1}{n}\sum([x, y] - \hat{\mu_u})^T([x, y] - \hat{\mu_u})$.

Figure 3(a) shows an example of GMM-Basic. Only one Gaussian distribution represents each user's check-in behavior. The user has 75 check-in records in the area with latitude between 30.3N and 30.6N, and longitude between 97.9W and 97.7W. Note that the x-axis is longitude and the y-axis is latitude. On the left side, GMM-Basic has higher probability density, but there is no check-in record there. Thus, only one Gaussian distribution could not describe each user's check-in behavior well.

GMM-Spatial: Each user's check-in records can be divided into several states, and each state can be modeled by one Gaussian distribution. Given the check-in records $\{(u, \ell = (x, y), t)\}$ of user u, u's mobility model is as follows:

$$p(\ell|u) = \sum_{i=1}^{K} p(\ell|Z_i, u)pr(Z_i|u), \tag{4}$$

[5] Although $p(\ell|u)$ equals $\int_t p(\ell|u,t)p(t)dt$, in practice, it is hard to derive $p(\ell|u)$ from the formula.

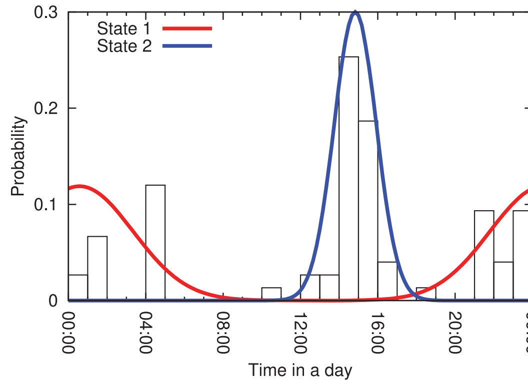


Fig. 4. Histogram of the check-in timestamp of a user in a day. The states correspond to Figure 3(c).

where K is the number of states, $p(\ell|Z_i)$ denotes the PDF of u checking in at ℓ in the state Z_i, and $pr(Z_i|u)$ denotes the probability of u in state Z_i. $pr(Z_i|u)$ and the parameters of $p(\ell|Z_i, u)$ can be learned by the EM algorithm.

As presented in Cho et al. [2011], K is set to 2 and these states represent the home and work states. Figure 3(b) shows an example of GMM-Spatial with two states. Two Gaussian distributions are used to represent each user's check-in behavior with two different states. Obviously, GMM-Spatial is better than GMM-Basic for describing each user's check-in behavior. However, each state could not be well represented by one bivariate Gaussian distribution such as state 2 in Figure 3(b).

GMM-Temporal: Since users may have different states of check-in behavior at different times, the wrapped Gaussian distribution[6] is selected to describe the work state and home state on the timeline [Cho et al. 2011]. Given the check-in records $\{(u, \ell = (x, y), t)\}$ of user u, u's mobility model is the same as GMM-Spatial (Equation (4)) and is also learned via the EM algorithm. However, in the E-step of the EM algorithm, the check-in records are classified by the Bayesian classifier on the timeline in one day. The PDF of a time t of state Z_i is as follows:

$$p(t|Z_i, u) = \frac{1}{\sigma_{u,i}\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left(\frac{-(t - \mu_{u,i} + 2\pi k)^2}{2(\sigma_{u,i})^2}\right),$$

where the time t is mapped onto $[0, 2\pi)$. Then, in the M-step, the parameters of $p(t|Z_i)$ will be updated. When the parameters converge, the check-in records of each state will be represented by one bivariate Gaussian distribution in the spatial domain.

Figure 3(c) shows an example of GMM-Temporal with two states, and Figure 4 shows the corresponding check-in time distribution in a day. The distribution is not obvious to distinguish into two bivariate Gaussian distributions since the check-in records are classified by the timestamp but not the coordinate. Thus, the two bivariate Gaussian distributions overlap. However, for each type of check-in record, it is also hard to describe the check-in records completely.

GMM-Category: GMM-Basic, GMM-Spatial, and GMM-Temporal are based only on the coordinate information of locations, i.e., $\ell = (x, y)$. To take the category into consideration, ℓ can be represented as $\ell = (x, y, c)$, where c is the category corresponding

[6]In Cho et al. [2011], the authors selected the truncated Gaussian distribution for the time distribution. However, the time of a day is circular. Thus, the wrapped Gaussian distribution is better than truncated Gaussian distribution for the time distribution here.

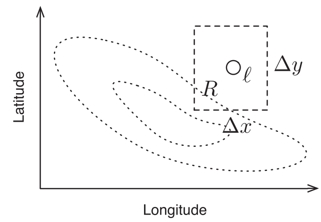


Fig. 5. An example of calculating the propagation probability in GMMs.

to ℓ.[7] Given the check-in records $\{(u, \ell = (x, y, c), t)\}$ of user u, u's mobility model is as follows:

$$p(\ell|u) = p(x, y, c|u) = p(x, y|c, u)pr(c|u), \tag{5}$$

where the $p(x, y, c|u)$ represents the spatial distribution of u under category c, i.e., the spatial preference of u under category c, and $pr(c|u)$ represents the probability of u checking in at a location that belongs to category c.

To derive $p(x, y|c, u)$, it is easy to utilize GMM-Basic, GMM-Spatial, or GMM-Temporal to represent the spatial distribution under category c since GMM-Basic, GMM-Spatial, and GMM-Temporal are only based on the spatial domain. However, the check-in records have the spatial temporal sparsity issue. A user may have few check-in records under a given category. For this situation, we set a global threshold θ_c for each category. Given a coordinate (x, y) and a category c, to derive $p(x, y|c, u)$, u's check-in records that belong to c are selected to build the Gaussian distribution if u has a sufficient number (i.e., larger than θ_c) of check-in records that belong to c; otherwise we utilize the check-in preference under category c in the region $R_{x,y,r}$, $p(x, y|c, R_{x,y,r})$, instead of directly calculating $p(x, y|c, u)$, where $R_{x,y,r}$ is a circular region with center (x, y) and radius r. For example, if GMM-Basic is adopted for $p(x, y|c, u)$, and then $p(x, y|c, u)$ can be represented as follows:

$$p(x, y|c, u) = \begin{cases} \mathcal{N}(\mu_{u,c}, \Sigma_{u,c}) \text{ if } |\{(u, \ell = (x, y, c), t)|u, c\}| > \theta_c \\ p(x, y|c, R_{x,y,r}) = \mathcal{N}(\mu_{R_{x,y,r},c}, \Sigma_{R_{x,y,r},c}) \text{ otherwise,} \end{cases} \tag{6}$$

where $\mu_{u,c}$ and $\Sigma_{u,c}$ are derived from u's check-in records under category c, $\{(u, \ell = (x, y, c), t)|c, u\}$. Moreover, $\mu_{R_{x,y,r},c}$ and $\Sigma_{R_{x,y,r},c}$ are derived from the check-in records under category c in the region $R_{x,y,r}$, $\{(u, \ell = (x, y, c), t)|c, (x, y) \in R_{x,y,r}\}$.

To derive $pr(c|u)$, we adopt logistic regression with regularization to estimate the probability. The 12 features are selected, number of check-ins, number of friends, max/min/avg of inter check-in time, max/min/avg of inter check-in distance, category entropy, and time entropy (hours of a day/days of week/weekday or weekend).

Deriving propagation probability: The GMMs have two-dimensional distribution. Thus, it is necessary to select a region to calculate the probability. Figure 5 shows an example of calculating the propagation probability in GMMs, where the target location ℓ, and the region R around ℓ are controlled by Δx and Δy. The probability of v checking in at ℓ is as follows:

$$pp(u, v|\ell) \approx pr(\ell|v) = \int_R p(\ell|v)\mathrm{d}A, \tag{7}$$

where $p(\ell|v)$ is from Equation (3) or (4). However, it is hard to set the size of Δx and Δy. Thus, we will show another approach to derive the propagation probability in an LBSN.

[7]In this article, we define that each location ℓ has only one corresponding category c.

Estimation error: The estimation error describes the error between the mobility model from observations and the real mobility model. Here, we select Mean Square Error (MSE) to describe the situation. The details are as follows:

$$\mathrm{MSE}(\hat{\mu}) = \mathrm{Var}(\hat{\mu}) = \frac{1}{n}[\sigma_x^2 + \sigma_y^2],$$

and

$$\begin{aligned}\mathrm{MSE}(\hat{\Sigma}) &= \mathrm{Var}(\hat{\Sigma}) = trace(\mathrm{Var}(S^2)) \\ &= \frac{2}{n-1}\sigma_x^2 + \frac{2}{n-1}\sigma_y^2 = \frac{2}{n-1}(\sigma_x^2 + \sigma_y^2).\end{aligned}$$

Thus, the MSE of $\hat{\mu}$ and $\hat{\Sigma}$ are $\propto \frac{1}{n}$, where n denotes the number of check-in records.

Complexity analysis: Suppose there is a set of u's check-in records (u, ℓ, t) with size n. To build $p(\ell|u)$ using Equation (3) for GMM-Basic, the time cost is $O(n)$. Moreover, to build $p(\ell|u)$ in Equation (4) for GMM-Spatial, the time cost of E-step and M-step is Kn and $O(Kn)$ if there are K states, respectively. Then, the total time cost is $I(Kn + O(Kn)) = O(IKn)$, where I is the number of iterations. Similar to GMM-Spatial, to build $p(\ell|u)$ in Equation (4) for GMM-Temporal, the time cost is $O(IKn)$. On the other hand, given a target ℓ, to derive $pp(u, v|\ell)$ via Equation (7), the time cost for GMM-Basic is $O(1)$. For GMM-Spatial and Temporal, the time cost is $K \cdot O(1) = O(K)$.

5.2. Distance-Based Approaches

In Section 5.1, we recognize that it is hard to calculate the propagation probability from GMMs. Moreover, using GMMs cannot capture the order of the check-in records. We argue that different orders of check-in records have different check-in behavior.[8] We think that the distance between the consecutive check-in records will be changed if the order of check-in records is changed. Therefore, we utilize not only the coordination of check-in records, but also the distance between the consecutive check-in records to describe individual check-in behavior. Based on this idea, we propose novel DMMs to capture individual check-in behavior.

DMM-Basic: This is the basic form of DMM. The idea of DMM is that it simulates a user staying at one location and then checking in at the target location. Therefore, DMM has two layers. The first layer is the stationary distribution of visited locations, and the second is the probability density of u checking in with the distance between visited locations and the target location from a visited location. Based on the concept above, the PDF of u checking in at a location ℓ, $p(\ell|u)$, can be shown as follows:

$$\begin{aligned}p(\ell|u) &= \sum_l pr(u \text{ stays at } l\text{, and then checks in at } \ell) \\ &= \sum_l pr(u \text{ is at } l) \cdot p(u \text{ checks in with distance } d(l, \ell) \text{ from } l) \\ &= \sum_l p_{u,l} f_u(d(l, \ell)),\end{aligned} \tag{8}$$

[8]Other temporal features are important for modeling check-in behavior, such as the timestamp of check-in records. In this article, we focus on evaluating the probability of a user checking in at a location. The actual time of a user checking in at a location is not important for the location promotion problem since we just care whether the user checks in at a location or not. Thus, the distance-based approach only adopts the order feature based on the timestamp of check-in records for modeling check-in behavior.

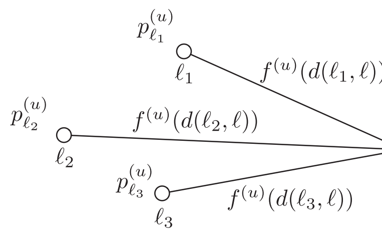


Fig. 6. The concept of the distance-based mobility model.

where $p_{u,l}$ denotes the stationary probability of user u at visited location l, $f_u(d(l,\ell)$ denotes the PDF of user u moving from a visited location l to the target location ℓ, and $d(l,\ell)$ denotes the distance between l and ℓ.

To derive the stationary distribution of visited locations, there are many approaches, such as random walk with restart (RWR) and lazy random walk (LRW). To select a better solution for the stationary distribution, we utilize an empirical approach. First, we compare RWR and LRW with different parameters in Section 6.4. Second, we think that the user's current location is close to the locations of recent check-in records. Thus, we can set a time interval before the current time such that only check-in records in the time interval are extracted to evaluate the stationary probability. We compare different lengths of time interval in Section 6.5.

Figure 6 shows an example of the DMM. The user u has three visited locations ℓ_1, ℓ_2, and ℓ_3. Assume that u has five check-in records (u,ℓ_1,t_1), (u,ℓ_2,t_2), (u,ℓ_3,t_3), (u,ℓ_1,t_4), and (u,ℓ_3,t_5), where $t_i < t_j$ if $i < j$. Thus, the transition probabilities are $pr(\ell_1 \to \ell_2) = 0.5$, $pr(\ell_1 \to \ell_3) = 0.5$, $pr(\ell_2 \to \ell_3) = 1$, and $pr(\ell_3 \to \ell_1) = 1$. If RWR is selected to derive the stationary probabilities, and then $p_{u,\ell_1} = 0.39$, $p_{u,\ell_2} = 0.21$ and $p_{u,\ell_3} = 0.40$, where the restart probability is 0.15 and the initial distribution of ℓ_1, ℓ_2, and ℓ_3 is uniform.

The Pareto distribution is selected for the distance distribution f_u since the next check-in distance in LBSNs has a self-similar property [Zhu et al. 2014]. The form of Pareto distribution is as follows:

$$f(x;\alpha,\beta) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}, \tag{9}$$

where α is the shape parameter and β is the minimum of x. Since β cannot be 0, we fix $\beta = 1$. Thus, based on Equations (8) and (9), DMM-Basic can be formulated as follows:

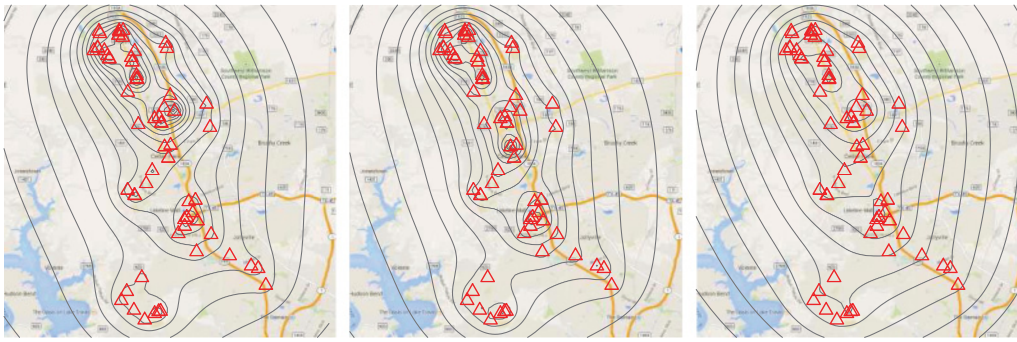
$$p(\ell|u) = \sum_{l} \frac{p_{u,l}\alpha_u}{(d(l,\ell)+1)^{\alpha_u+1}}, \tag{10}$$

where α_u denotes the distance preference of u. Given a sequence of u's check-in records $\langle(u,\ell_1,t_1), (u,\ell_2,t_2), \ldots, (u,\ell_n,t_n)\rangle$, we can derive the distances between consecutive check-in records $x_1, x_2, \ldots,$ where $x_{n-1}, x_i = d(\ell_i,\ell_{i+1})$, $i = 1,2,\ldots,n-1$. Based on MLE, the estimator of α_u is as follows:

$$\hat{\alpha_u} = \frac{n-1}{\sum_{i=1}^{n-1} \ln(x_i+1)}. \tag{11}$$

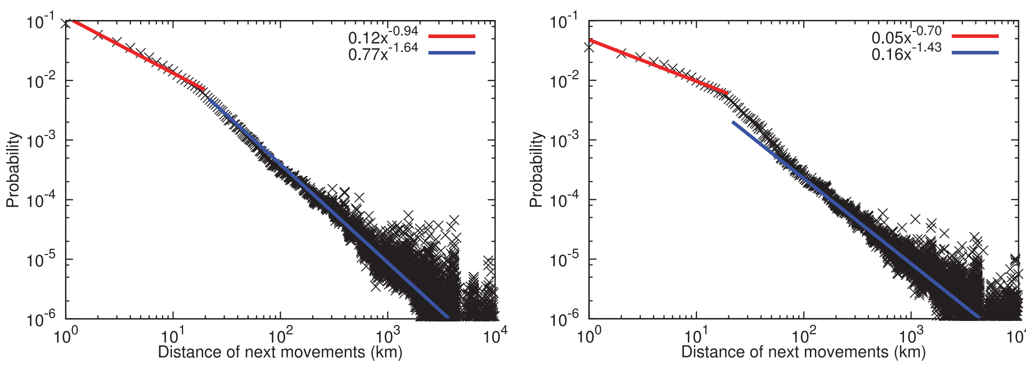
If u's next check-in distances are far, α_u is lower. Then, $p(\ell|u)$ will be higher since u is likely to check in at a location far away from the location of u's last check-in records and vice versa. This shows that DMM is able to reflect individual distance preferences.

Figure 7(a) shows an example of DMM-Basic. DMM-Basic is more representative than GMMs for individual check-in behavior. If we modify the order of the check-in



(a) DMM-Basic (b) DMM-Basic (different order) (c) DMM-Social

Fig. 7. Examples of distance-based mobility models, where the user is the same as in Figure 3.



(a) gowalla (b) brightkite

Fig. 8. The distribution of distance of next movement in two different datasets.

records, DMM shows different results (Figure 7(b)). DMM is different from GMMs since GMMs focus on the coordinates of the check-in records but not on the relations between them. There are two differences if the order is modified. First, the distribution of higher probability density in Figure 7(a) is different from the distribution of higher probability density in Figure 7(b). This shows that the stationary distribution is changed if the order is modified. Second, the distribution probability density in Figure 7(b) is smoother than in Figure 7(a). This shows that the user is more likely to check in at distant locations.

DMM-Social: Users' check-in behavior can be influenced by their friends in LBSNs, and it can be reflected by the next check-in distance [Ye et al. 2011; Cho et al. 2011]. Therefore, the next check-in distances are divided into two parts, locality behavior and social influenced behavior. Locality behavior means that the check-in record is based on their locality behavior, and social influenced behavior means that the check-in record is influenced by their friends. Figure 8 shows the distribution of distance of next movement in two different datasets. The distributions can be divided into two parts. Similar observations are also found in Cho et al. [2011]. DMM-Social can be shown as follows:

$$p(\ell|u) = \sum_{l} p_{u,l}[f_{u,\mathrm{M}}(d(l,\ell))pr(\mathrm{M}|u) + f_{u,\mathrm{S}}(d(l,\ell))pr(\mathrm{S}|u)], \tag{12}$$

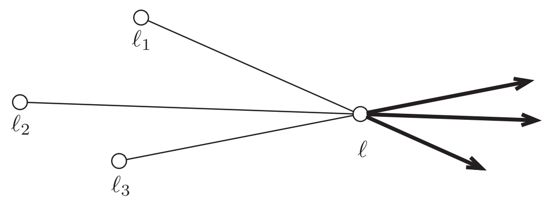


Fig. 9. An example of calculating the propagation probability in DMMs.

where $pr(\mathrm{M}|u)$ and $pr(\mathrm{S}|u)$ denote the probability of a distance contributed by locality and social influence, respectively. Moreover, $f_{u,\mathrm{M}}(d(l,\ell))$ and $f_{u,\mathrm{S}}(d(l,\ell))$ denote the probability density of check-in $d(l,\ell)$ controlled by locality and social influence, respectively. For DMM, the Pareto distribution is also utilized for $f_{u,\mathrm{M}}$ and $f_{u,\mathrm{S}}$. Thus, based on Equation (9), Equation (12) can be rewritten as follows:

$$p(\ell|u) = \sum_{l} p_{u,l}\left[\frac{\alpha_{u,\mathrm{M}} \cdot pr(\mathrm{M}|u)}{(d(l,\ell)+1)^{\alpha_{u,\mathrm{M}}+1}} + \frac{\alpha_{u,\mathrm{S}} \cdot pr(\mathrm{S}|u)}{(d(l,\ell)+1)^{\alpha_{u,\mathrm{S}}+1}}\right], \tag{13}$$

where $pr(\mathrm{M}|u)$, $pr(\mathrm{S}|u)$, $\alpha_{u,\mathrm{M}}$, and $\alpha_{u,\mathrm{S}}$ can be learned by the EM algorithm.

Figure 7(c) shows an example of DMM-Social. DMM-Social is smoother than the DMM-Basic in Figures 7(a) and (b). This reflects the next check-in distance that will be influenced by friends in Figure 8. Moreover, the coordinate of check-in records is also considered by DMMs.

DMM-Category: Similar to GMM-Category, given the check-in records $\{(u, \ell = (x, y, c), t)\}$ of user u, u's mobility model is the same as in Equation (5). The $p(x, y|c, u)$ can be derived from DMM-Basic or DMM-Social. If DMM-Basic is adopted for $p(x, y|c, u)$, $p(x, y|c, u)$ can be represented as follows:

$$p(x, y|c, u) = \begin{cases} \sum_l \frac{p_{u,l}\alpha_{u,c}}{(d(l,\ell)+1)^{\alpha_{u,c}+1}} & \text{if } |\{(u, \ell = (x, y, c), t)|u, c\}| > \theta_c \\ p(x, y|c, R_{x,y,r}) = \sum_l \frac{p_{u,l}\alpha_{R_{x,y,r},c}}{(d(l,\ell)+1)^{\alpha_{R_{x,y,r},c}+1}} & \text{otherwise}, \end{cases} \tag{14}$$

where $\alpha_{u,c}$ is derived from u's check-in records under category c, $\{(u, \ell = (x, y, c), t)|c, u\}$. Moreover, $\alpha_{R_{x,y,r},c}$ is derived from the check-in records under category c in the region $R_{x,y,r}$, $\{(u, \ell = (x, y, c), t)|c, (x, y) \in R_{x,y,r}\}$.

Deriving propagation probability: If a user tends to check in at a distance far from both the user's current location and the target location, there is a higher chance that the user will check in at the target location. Thus, to calculate the probability based on a DMM, we just calculate the probability of check-in with the distance that exceeds the distance between the visited locations and the target location. Figure 9 shows an example of calculating probability based on DMMs. Given the target location ℓ and visited locations ℓ_1, ℓ_2, and ℓ_3, it only accumulates the probability of the user's moving distance exceeding the distance from each visited location to the target location (bold segments). Thus, the probability of v checking in at ℓ is as follows:

$$pr(\ell|v) = \sum_{l} p_{u,l}\int_{d(l,\ell)+1}^{\infty} f_v(x)\mathrm{d}x. \tag{15}$$

Furthermore, in DMM-Basic, $pp(u, v|\ell)$ is as follows:

$$pp(u, v|\ell) \approx \sum_{l} \frac{p_{v,l}}{(d(l,\ell)+1)^{\alpha_v}}. \tag{16}$$

Then, in DMM-Social, $pp(u, v|\ell)$ is as follows:

$$pp(u, v|\ell) \approx \sum_{l} p_{v,l} \left[\frac{pr(\mathrm{M}|v)}{(d(l, \ell) + 1)^{\alpha_{v,\mathrm{M}}}} + \frac{pr(\mathrm{S}|v)}{(d(l, \ell) + 1)^{\alpha_{v,\mathrm{S}}}} \right]. \tag{17}$$

Estimation error: Given a set of u's check-in records $\{(u, \ell, t)\}$ with size $n + 1$, the estimation error of the DMMs is as follows:

$$\begin{aligned} \mathrm{MSE}(\hat{\alpha_u}) = \mathrm{Var}(\hat{\alpha_u}) &= \mathrm{Var}\left(\frac{n}{\sum_i \ln(x_i + 1)} \right) \\ &= n^2 \,\mathrm{Var}\left(\frac{1}{\sum_i y_i} \right), \text{ where } y = \ln(x_i + 1) \\ &\approx n^2 \frac{\alpha_u^4}{n^3} = \frac{\alpha_u^4}{n}. \end{aligned}$$

The MSE of $\hat{\alpha}$ is $\propto \frac{1}{n}$. The results[9] show that the estimation error of the DMMs is similar to the estimation error of the GMMs.

Complexity analysis: Suppose there is a set of u's check-in records $\{(u, \ell, t)\}$ with size n. To build $p(\ell|u)$ using Equation (10) for DMM-Basic, first the time cost of calculating the next check-in distances is n. Then, to calculate $\hat{\alpha_u}$ using Equation (11), the time cost is n. Thus, the time cost of building DMM-Basic is $n + n = O(n)$. Moreover, to build $p(\ell|u)$ in Equation (13) for DMM-Social, the time cost of E-step and M-step is $2n$ and $O(2n)$, respectively. Then, the total time cost is $I(2n + O(2n)) = O(In)$, where I is the number of iterations. On the other hand, given a target ℓ, to derive $pp(u, v|\ell)$ via Equation (16), the time cost for DMM-Basic is $O(n)$. For DMM-Social via Equation (17), the time cost is $2 \cdot O(n) = O(n)$.

6. PERFORMANCE EVALUATION

In this section, extensive experiments are conducted to evaluate the effectiveness of our proposed mobility models, GMMs and DMMs. Moreover, we also show the results of static propagation probability and location-aware propagation probability in LBSNs. We implemented the proposed models in Python.

6.1. Setting

6.1.1. Datasets Description. In this article, we have selected the gowalla and brightkite datasets [Cho et al. 2011] for observation and evaluation.[10] There were 196,591 users, 950,327 social connections, and 6,442,890 check-ins during February 2009–October 2010 in the gowalla dataset, and there are 58,228 users, 214,078 social connections and 4,491,143 check-ins during April 2008–October 2010 in the brightkite dataset.

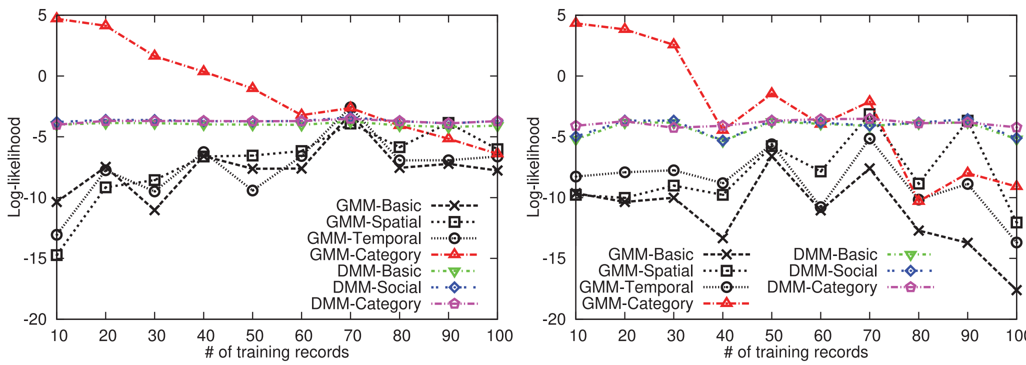
6.1.2. Measurements. In this article, for the location promotion problem, our approach is to propose mobility models to capture individual check-in behavior, and then utilize the proposed mobility models to set the propagation probability in an LBSN. Thus, in the experiments, to evaluate different mobility models and different approaches for propagation probability, the measurements, log-likelihood, and Receiver operating characteristic curve are selected, respectively. The detailed descriptions of the two measurements are as follows.

Log-likelihood: To compare the performance of different mobility models, log-likelihood is selected to measure the mobility models for LBSNs since the mobility model is a two-dimensional probability density function [Cho et al. 2011; Gao et al.

[9]The approximation is based on the delta method.
[10]Both of these datasets are public datasets that can be downloaded from https://snap.stanford.edu/data/.

Table I. The Number of Users and The Different Number of Training Records

Datasets	# of training records 10	20	30	40	50	60	70	80	90	100	Total users
gowalla	1,568	10,648	444	4,604	244	2,560	137	1,614	94	1,079	22,992
brightkite	636	315	215	143	111	80	62	64	40	31	1,697



(a) gowalla (b) brightkite

Fig. 10. The results of GMMs and DMMs with different numbers of training records in different LBSNs.

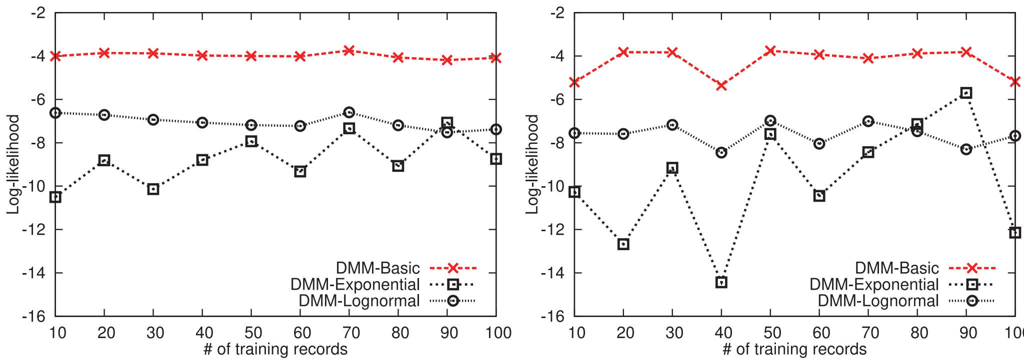
2013b; Lichman and Smyth 2014]. If log-likelihood is higher, the mobility model represents the real check-in behavior of each user well.

ROC curve: To compare different approaches for propagation probability, the ROC curve is selected to show the results of different global activation thresholds [Goyal et al. 2010; Barbieri et al. 2012].

6.2. Comparison of Different Mobility Models

Due to the spatial and temporal sparsity issues of check-in records, most users seldom perform check-in (10–100 records). Therefore, we observe the relation between the number of training data and log-likelihood of each mobility model here. The proportion of training and testing of check-in records is 80% and 20%, and the number of training records is 10, 20, ..., 100 in the two datasets. The details of the number of users are given in Table I. Furthermore, the comparisons are `GMM-Basic`, `GMM-Spatial`, `GMM-Temporal`, `GMM-Category`, `DMM-Basic`, `DMM-Social`, and `DMM-Category`. For `GMM-Category` and `DMM-Category`, the range r and threshold θ_c is set to 1km and 10, respectively.

Figure 10 shows the results of comparison with different numbers of training records in the two datasets. First, we observe the results of the methods that only consider the spatial-temporal features of check-in records. In the gowalla dataset, DMMs present higher log-likelihood value than GMMs. Since the state-of-the-art approaches [Cho et al. 2011; Gao et al. 2013b; Lichman and Smyth 2014] proposed GMMs, our proposed GMMs further consider spatial and temporal information. As shown in Figure 10, DMMs can better reflect individual check-in behavior than GMMs, which demonstrates the advantage of DMMs. However, when the number of training data is 70, all approaches have higher log-likelihood value. It can be considered as a bias since the number of data that contain 70 training records is lower than others (Table I). On the other hand, `DMM-Social` has higher log-likelihood value than `DMM-Basic`. By exploring the social affection in distance preference in LBSNs, `DMM-Social` is closer to real individual check-in behavior. Furthermore, DMMs can capture individual check-in behavior by only 10 training records. This shows that DMMs deal with the spatial and temporal sparsity issues of check-in records. The results in brightkite are similar to



(a) gowalla (b) brightkite

Fig. 11. The results of three distance distributions in `DMM-Basic` with different numbers of training records in different LBSNs.

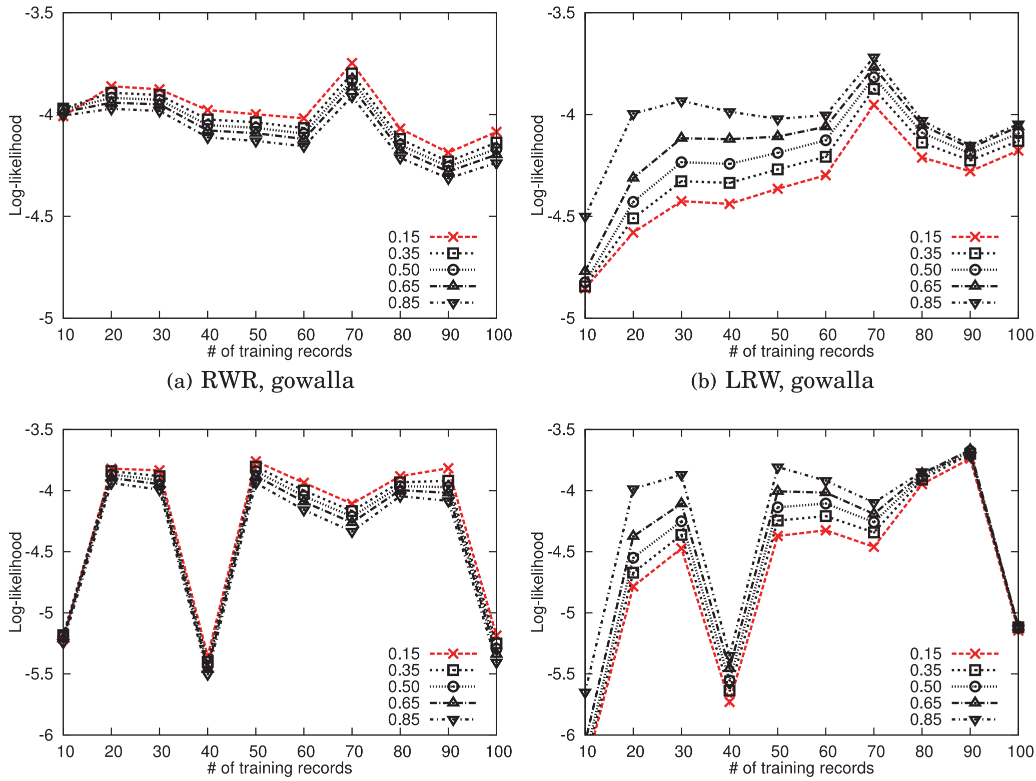
the results in gowalla. However, the amplitude of GMMs is large since the number of data we selected is lower.

Next, we observe the results category-based approaches. The performance of `DMM-Category` is similar to the performance of `DMM-Social` in the two datasets. In substance, `DMM-Category` has higher log-likelihood than `DMM-Social`, but it is insignificant. Especially, `GMM-Category` has abnormally high log-likelihood when the number of training records is 10, and the log-likelihood falls when the number of training records is larger in the two datasets. The reason is that the probability of `GMM-Category` adopting the global category preference is higher when the number of training records is lower. When the number of training records is 10, `GMM-Category` almost adopts the global category preference since the threshold θ_c is 10. In `GMM-Category`, the global category preference only considers the check-in coordinates in the area around the target location. Therefore, the probability density is gathered in this area. Moreover, `GMM-Category` only adopts the global category preference when a user does not have enough check-in records in a given category. Hence, `GMM-Category` has abnormally high log-likelihood when the number of training records is 10. When the number of training records is larger, the probability of `GMM-Category` adopting global category preference is lower. Thus, the log-likelihood of `GMM-Category` is lower when the number of training records is increasing. On the other hand, when a user does not have enough check-in records in a given category, `DMM-Category` adopts not only global category preference (distance preference), but also the user visited coordinations. Therefore, `DMM-Category` performs more stably than `GMM-Category`.

6.3. Comparison of Different Distributions for Distance

In DMMs, the Pareto distribution is selected to represent the distance of the next movement of each user. Some related distributions (such as exponential and log-normal distribution) are listed. The domain of these two distributions is $[0, \infty)$, and they are related to Pareto and Gaussian distributions, respectively. To compare different distance distributions in DMMs, `DMM-Basic` is selected since it is the simplest form of the DMMs. Then, the distance distribution in `DMM-Basic` is replaced with exponential and log-normal distribution, respectively. The modified `DMM-Basic` is called `DMM-Exponential` and `DMM-Lognormal`, respectively.

Figure 11 shows the results of the comparison of `DMM-Basic`, `DMM-Exponential`, and `DMM-Lognormal`. `DMM-Basic` has higher log-likelihood value in the two datasets. `DMM-Exponential` has the lowest log-likelihood value such that the distance of the next



(a) RWR, gowalla
(b) LRW, gowalla
(c) RWR, brightkite
(d) LRW, brightkite

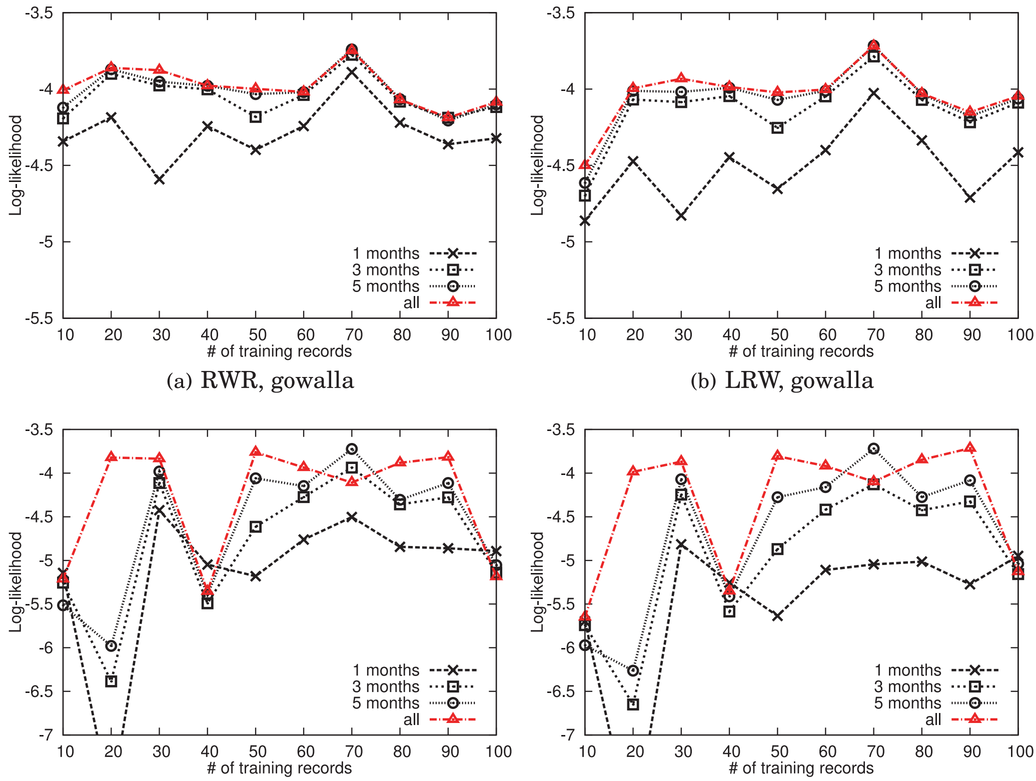
Fig. 12. The results of different random walk models for the stationary distribution in DMM-Basic with different numbers of training records in different LBSNs.

movement is not memoryless. Finally, the results show that the Pareto distribution is more suitable for the distance of next movements in LBSNs than the exponential and log-normal distribution in both datasets.

6.4. Comparison of Different Random Walk Models

In Section 5.2, DMMs consist of the stationary distribution and the distance distribution, and Section 6.3 shows the results of different distance distributions. In this section, we aim to find a suitable solution via an empirical study from different random walk models for the stationary distribution in DMMs. RWR and LRW are selected to derive the stationary distribution with different restart probabilities and staying probabilities, respectively. The distribution of the initial vector is uniform, and the restart probability and staying probability are set to 0.15, 0.35, 0.5, 0.65, and 0.85.

Figure 12 shows the results of `DMM-basic` using RWR and LRW with different probabilities in two datasets. In Figure 12(a), RWR with the restart probability 0.15 (abbreviated as RWR 0.15) has the highest log-likelihood in the gowalla dataset. Note that RWR 0.15 is PageRank [Han et al. 2011], the most common setting of RWR. Moreover, in Figure 12(b), LRW with staying probability 0.85 (abbreviated as LRW 0.85) has the highest log-likelihood in the gowalla dataset. The log-likelihood is higher when RWR has higher restart probability, whereas the log-likelihood is higher when LRW has higher staying probability. The results show that RWR and LRW have different characteristics for stationary distribution.



(a) RWR, gowalla (b) LRW, gowalla

(c) RWR, brightkite (d) LRW, brightkite

Fig. 13. The results of different window sizes for the stationary distribution in `DMM-Basic` with different numbers of training records in different LBSNs.

In Figures 12(a) and (b), the staying probability of LRW is more sensitive than the restart probability of RWR, especially when the number of training records is lower. Furthermore, the results show that RWR 0.15 is similar to LRW 0.85, but RWR 0.15 has higher log-likelihood when the number of training records is lower. The results in the brightkite dataset (Figures 12(c) and (d)) are similar to the results in the gowalla dataset. Based on the above results, we select RWR 0.15 as the default setting for the stationary distribution in DMMs.

6.5. Impact of Window Size

As mentioned in Section 5.2, we guess that a user's current location is near the locations of recent check-in records. Thus, in DMMs, we consider only recently visited locations from check-in records for the stationary distribution. Given u's check-in records $\{(u, \ell, t)\}$, we set a time interval $[ct - \Delta t, ct]$, where ct is the current time and Δt is the window size. $\{(u, \ell, t) | t \in [ct - \Delta t, ct]\}$. All u's check-in records that are in $[ct - \Delta t, ct]$ are used to derive the stationary distribution. In this section, we observe the impact of `DMM-basic` with different window sizes, and then we can select a suitable window size for DMMs. The window size is set to 1, 3, 5 months, and ∞ (considering all check-in records), and the current time is set to the time of the last training record.

Figure 13 shows the results of window sizes for the stationary distribution in `DMM-Basic` with different numbers of training records in the two datasets. In Figures 13(a) and (b), the results show that the log-likelihood is higher when the window size of

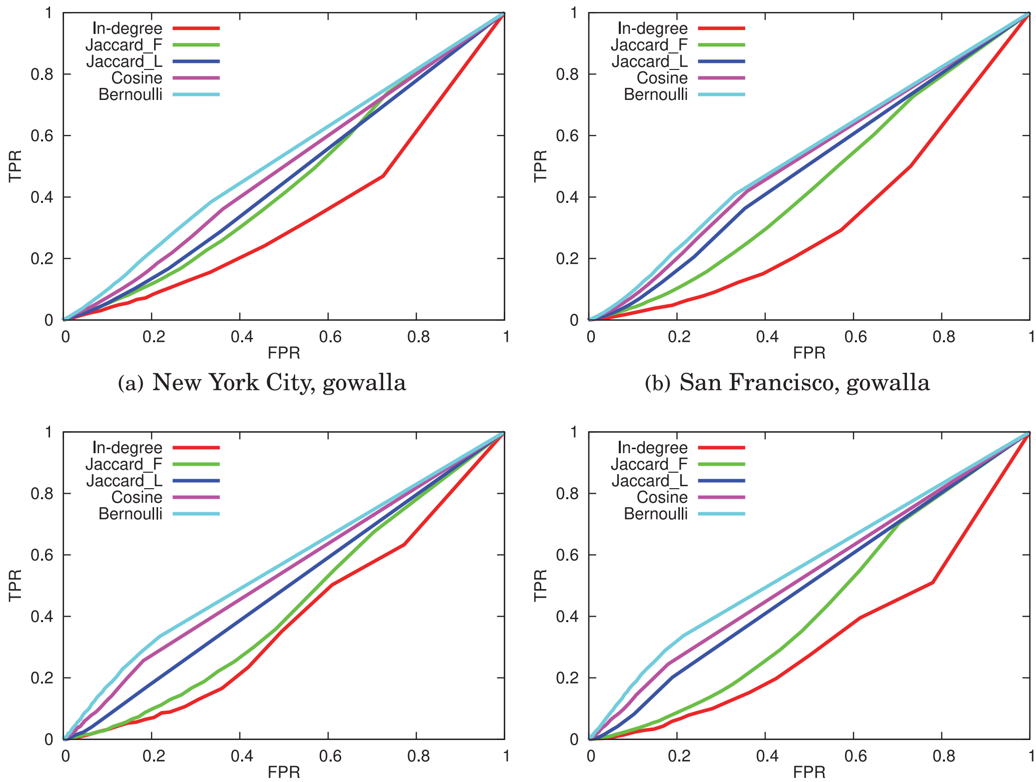
RWR and LRW is higher in the gowalla dataset. Moreover, the window size of LRW is more sensitive than the window size of RWR. Figures 13(c) and (d) show similar results in the brightkite dataset. The results are the opposite of what we expected. We think there are two possible reasons. First, users are more likely to check in in the vicinity of visited locations [Cho et al. 2011]. Thus, considering more check-in records can help to increase the performance of DMMs. Second, if we consider only recent check-in records for stationary distribution, it reduces the number of training records, and the sparsity issue of check-in records affects the performance of DMMs. On the other hand, when the window size is larger than 3 months, the results are similar in the gowalla dataset, but not in the brightkite dataset. We think that this is due to the different check-in preferences in different LBSNs. According to the results and above discussions, the default setting of the window size in DMMs is ∞.

6.6. Comparison of Methods for Propagation Probability

To observe the information propagation in an LBSN, two target locations are selected: one is San Francisco Caltrain Station, San Francisco (37.776430N 122.394318W), and the other is Central Park, New York City (40.780606N 73.968088W). Moreover, we only selected users who have 10 check-in records or above. There are 67,653 users and 629,031 edges selected in the gowalla dataset, and 24,100 users and 243,922 edges selected in the brightkite dataset. To determine whether a user is active or not, we set the active range of the target location. Note that the active range is also used to calculate the propagation probability of GMMs. If one of the user's check-in records is in the active range, the user is active in the ground truth. In the experiments, the radius of the active range is set to 500m. For the approaches in static propagation probability, five approaches are selected in the experiments, where `In-degree` denotes the in-degree of nodes, `Jaccard_F` denotes the Jaccard index of friends, `Jaccard_L` denotes the Jaccard index of locations, `Cosine` denotes cosine of locations, and `Bernoulli` denotes the Bernoulli estimator.

Figure 14 shows the results of five approaches for static propagation probability. In the gowalla dataset (Figures 14(a) and (b)), `Bernoulli` performs better than the other approaches of static probability, but not significantly. Although `Bernoulli` learns from history records, it cannot adapt to the information propagation in LBSNs. The ROC curves of `Jaccard_F`, `Jaccard_L` and `Cosine` are closest to the diagonal. Thus, the number of common friends and locations is not related to the information propagation in LBSNs. Furthermore, `In-degree`, the most commonly used method in social networks [Kempe et al. 2003; Chen et al. 2010; Li et al. 2014] is worse than random guess. Thus, the in-degree of nodes and the degree of information propagation in LBSNs are not helpful in the location promotion problem. Finally, the results show that the ROC curves of the five approaches are close to the diagonal in both target locations. From the above experimental results, the static propagation probability is not able to truly reflect the information propagation in LBSNs. The results in the brightkite dataset are shown in Figures 14(c) and (d), and these results depict a similar phenomenon.

For location-aware propagation probability, seven approaches are selected, `GMM-Basic`, `GMM-Spatial`, `GMM-Temporal`, `GMM-Category`, `DMM-Basic`, `DMM-Social`, and `DMM-Category`. For `GMM-Category` and `DMM-Category`, the category of two target locations, San Francisco Caltrain Station and Central Park, is set to "Travel & Transport" and "Outdoors & Recreation," respectively. Moreover, the range r and threshold θ_c is set to 1km and 10, respectively. Figure 15 shows the results of these location-aware approaches. First, we observe the results of the methods that only consider the spatial-temporal features of check-in records. In the gowalla dataset (Figures 15(a) and (b)), `DMM-Basic` has the largest Area Under Curve (AUC), where the target location is set to NYC. On the other hand, `DMM-Basic` and `DMM-Social` have similar AUC when the



(a) New York City, gowalla
(b) San Francisco, gowalla
(c) New York City, brightkite
(d) San Francisco, brightkite

Fig. 14. The results of five approaches for static propagation probability in different LBSNs.

target location is set to San Francisco. Both `DMM-Basic` and `DMM-Social` have better performance with different target locations. Moreover, GMMs have the lowest AUC and these curves of GMMs are close to the diagonal, where the diagonal curve represents the results of random guess. The reason is that the location-aware propagation probability from GMMs is controlled by the area around the target location. If the area is large, then the probability is higher. This reflects the disadvantage that it is hard to derive the location-aware propagation from GMMs. Particularly, `GMM-Spatial` has the best performance of the GMMs. This shows that individual check-in behavior can be divided into many states by spatial information. However, `GMM-Spatial` is limited by the disadvantage of calculating location-aware propagation of GMMs. As such, `GMM-Spatial` has lower AUC than DMMs. Finally, DMMs have larger AUC than GMMs, which indicates that DMMs are suitable to reflect the information propagation and derive the location-aware propagation in LBSNs. The results shown in Figures 15(c) and (d) are experiments in the brightkite dataset, which indicate similar results to those in the gowalla dataset.

Finally, we observe the results of the extension of GMM and DMM, `GMM-Category` and `DMM-Category`. In Figure 15, the results show that `GMM-Category` does not improve the performance by considering the category information of locations in the two datasets, whereas `DMM-Category` has significant performance improvement when the target location is set to NYC in the two datasets. Although `GMM-Category` and `DMM-Category` both consider the category information, `GMM-Category` does not have significant

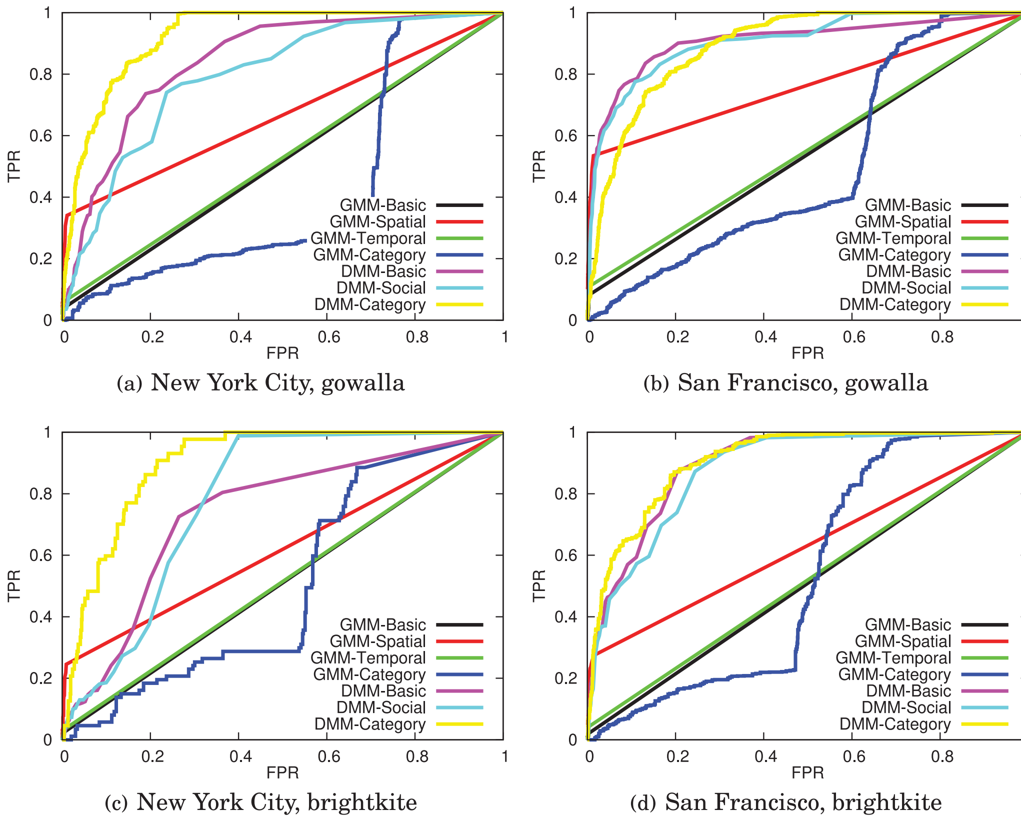


Fig. 15. The results of five approaches for location-aware propagation probability in different LBSNs.

performance improvement. For the models of GMMs and DMMs, GMMs only capture the coordinates of the check-in records, whereas DMMs consider not only the coordinates, but also the distances in the check-in records. When a user does not have enough check-in records in a given category, `GMM-Category` and `DMM-Category` both adopt global check-in behavior in the given category around the target location. The results (Figures 10 and 15) indicate that we cannot only consider global preference to infer the personal check-in behavior. However, due to the physical limitation of GMMs (only considering the coordinates), it is hard to integrate the global preference and personal preference. In contrast, with DMMs, it is easy to integrate the global preference (distance) and personal preference (coordinate). Moreover, due to the sparsity issue of check-in records, the situation of adopting in `GMM-Category` and `DMM-Category` often appears. Therefore, the distance preference is a key reason for `DMM-Category` outperforming `GMM-Category`.

7. CONCLUSION

In this article, we address the location promotion problem in LBSNs. Explicitly, the location promotion problem is formulated as an influence maximization problem. Given a target location and an LBSN, we aimed at deriving a set of k seed users to maximize the number of influenced users to check in at the target location. To describe the information propagation, we propose LICM based on ICM. However, the most challenging aspect is to derive the propagation probability with different target locations in LBSNs.

By referring to prior works on influence maximization, we developed some baselines to determine the propagation probabilities. Note that we claim that user mobility should be considered in deriving propagation probabilities in LBSNs. Users should check in at the target location and thus will spread the target location information to their friends. Since the information propagation is triggered by check in in LBSNs, we propose GMMs and DMMs, to capture individual check-in behavior in LBSNs. Moreover, we also extend GMMs and DMMs by taking into account the category information of locations. The experimental results show that DMMs perform better than other state-of-the-art approaches in terms of capturing individual check-in behavior and dealing with the spatial and temporal sparsity issues of the check-in data since it considers not only the coordinates of check-in records, but also the relations between check-in records. Moreover, DMMs really reflect the dynamic propagation probability with different target locations in LBSNs.

REFERENCES

Çigdem Aslay, Nicola Barbieri, Francesco Bonchi, and Ricardo A. Baeza-Yates. 2014. Online topic-aware influence maximization queries. In *EDBT*.

Jie Bao, Yu Zheng, and Mohamed F. Mokbel. 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In *ACM GIS*.

Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2012. Topic-aware social influence propagation models. In *IEEE ICDM*.

Panagiotis Bouros, Dimitris Sacharidis, and Nikos Bikakis. 2014. Regionally influential users in location-aware social networks. In *ACM GIS*.

Shuo Chen, Ju Fan, Guoliang Li, Jianhua Feng, Kian-lee Tan, and Jinhui Tang. 2015. Online topic-aware influence maximization. In *VLDB*.

Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *ACM KDD*.

Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *ACM KDD*.

Chen Cheng, Haiqin Yang, Irwin King, and Michael R. Lyu. 2012. Fused matrix factorization with geographical and social influence in location-based social networks. In *AAAI*.

Chen Cheng, Haiqin Yang, Michael R. Lyu, and Irwin King. 2013. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*.

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: User movement in location-based social networks. In *ACM KDD*.

Edith Cohen, Daniel Delling, Thomas Pajor, and Renato F. Werneck. 2014. Sketch-based influence maximization and computation: Scaling up with guarantees. In *ACM CIKM*.

Gregory Ference, Mao Ye, and Wang-Chien Lee. 2013. Location recommendation for out-of-town users in location-based social networks. In *ACM CIKM*.

Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2013a. Exploring temporal effects for location recommendation on location-based social networks. In *ACM RecSys*.

Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2013b. Modeling temporal effects of human mobile behavior on location-based social networks. In *ACM CIKM*.

Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2015. Content-aware point of interest recommendation on location-based social networks. In *AAAI*.

Fosca Giannotti, Mirco Nanni, and Dino Pedreschi. 2006. Efficient mining of temporally annotated sequences. In *SDM*.

Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. 2007. Trajectory pattern mining. In *ACM KDD*.

Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.

Amit Goyal, Francesco Bonchi, and Laks V. S. Lakshmanan. 2010. Learning influence probabilities in social networks. In *ACM WSDM*.

Amit Goyal, Francesco Bonchi, and Laks V. S. Lakshmanan. 2012. A data-based approach to social influence maximization. In *VLDB*.

Amit Goyal, Wei Lu, and Laks V. S. Lakshmanan. 2011. CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In *WWW*.

Jiawei Han, Micheline Kamber, and Jian Pei. 2011. *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Bo Hu and Martin Ester. 2013. Spatial topic modeling in online social media for location recommendation. In *ACM RecSys*.

Bo Hu and Martin Ester. 2014. Graph-based point-of-interest recommendation with geographical and temporal influences. In *ACM CIKM*.

Kyomin Jung, Wooram Heo, and Wei Chen. 2012. IRIE: Scalable and robust influence maximization in social networks. In *IEEE ICDM*.

Panos Kalnis, Nikos Mamoulis, and Spiridon Bakiras. 2005. On discovering moving clusters in spatio-temporal data. In *SSTD*.

David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *ACM KDD*.

David Kempe, Jon Kleinberg, and Éva Tardos. 2005. Influential nodes in a diffusion model for social networks. In *ICALP*.

Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *ACM KDD*.

Guoliang Li, Shuo Chen, Jianhua Feng, Kian-lee Tan, and Wen-Syan Li. 2014. Efficient location-aware influence maximization. In *ACM SIGMOD*.

Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *ACM KDD*.

Moshe Lichman and Padhraic Smyth. 2014. Modeling human location data with mixtures of kernel densities. In *ACM KDD*.

Cindy Xide Lin, Qiaozhu Mei, Jiawei Han, Yunliang Jiang, and Marina Danilevsky. 2011. The joint inference of topic diffusion and evolution in social communities. In *IEEE ICDM*.

Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. 2013. Learning geographical preferences for point-of-interest recommendation. In *ACM KDD*.

Bin Liu and Hui Xiong. 2013. Point-of-interest recommendation in location based social networks with topic and location awareness. In *SDM*.

Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang. 2010. Mining topic-level influence in heterogeneous networks. In *ACM CIKM*.

Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. A random walk around the city: New venue recommendation in location-based social networks. In *ASE/IEEE SocialCom*.

Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011. An empirical study of geographic user activity patterns in foursquare. In *AAAI ICWSM*.

Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. 2008. Prediction of information diffusion probabilities for independent cascade model. In *KES*.

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social influence analysis in large-scale networks. In *ACM KDD*.

Weiqing Wang, Hongzhi Yin, Ling Chen, Yizhou Sun, Shazia Sadiq, and Xiaofang Zhou. 2015. Geo-SAGE: A geographical sparse additive generative model for spatial item recommendation. In *ACM KDD*.

Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. TwitterRank: Finding topic-sensitive influential twitterers. In *ACM WSDM*.

Hao-Hsiang Wu and Mi-Yen Yeh. 2013. Influential nodes in one-wave diffusion model for location-based social networks. In *PAKDD*.

Jihang Ye, Zhe Zhu, and Hong Cheng. 2013. What's your next move: User activity prediction in location-based social networks. In *SDM*.

Mao Ye, Xingjie Liu, and Wang-Chien Lee. 2012. Exploring social influence for recommendation: A generative model approach. In *ACM SIGIR*.

Mao Ye, Peifeng Yin, and Wang-Chien Lee. 2010. Location recommendation for location-based social networks. In *ACM GIS*.

Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *ACM SIGIR*.

Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. 2013. LCARS: A location-content-aware recommender system. In *ACM KDD*.

Josh Jia-Ching Ying, Eric Hsueh-Chan Lu, Wang-Chien Lee, Tz-Chiao Weng, and Vincent S. Tseng. 2010. Mining user similarity from semantic trajectories. In *ACM LBSN*.

Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat-Thalmann. 2013. Time-aware point-of-interest recommendation. In *ACM SIGIR*.

Kai Zheng, Yu Zheng, Nicholas Jing Yuan, Shuo Shang, and Xiaofang Zhou. 2014. Online discovery of gathering patterns over trajectories. *IEEE Transactions on Knowledge and Data Engineering* 26, 8 (2014), 1974–1988.

Tao Zhou, Jiuxin Cao, Bo Liu, Shuai Xu, Ziqing Zhu, and Junzhou Luo. 2015. Location-based influence maximization in social networks. In *ACM CIKM*.

Wen-Yuan Zhu, Wen-Chih Peng, and Ling-Jyh Chen. 2014. Exploiting mobility for location promotion in location-based social networks. In *IEEE DSAA*.

Wen-Yuan Zhu, Wen-Chih Peng, Ling-Jyh Chen, Kai Zheng, and Xiaofang Zhou. 2015. Modeling user mobility for location promotion in location-based social networks. In *ACM KDD*.

Received November 2015; revised September 2016; accepted September 2016