

Article

Finding Potential Propagators and Customers in Location-Based Social Networks: An Embedding-Based Approach

Yi-Chun Chen and Cheng-Te Li * 

Institute of Data Science, National Cheng Kung University, Tainan City 70101, Taiwan; guava830420@gmail.com

* Correspondence: chengte@mail.ncku.edu.tw; Tel.: +886-6-2757575 (ext. 53628)

Received: 23 October 2020; Accepted: 10 November 2020; Published: 11 November 2020



Abstract: In the scenarios of location-based social networks (LBSN), the goal of location promotion is to find information propagators to promote a specific point-of-interest (POI). While existing studies mainly focus on accurately recommending POIs for users, less effort is made for identifying propagators in LBSN. In this work, we propose and tackle two novel tasks, *Targeted Propagator Discovery* (TPD) and *Targeted Customer Discovery* (TCD), in the context of Location Promotion. Given a target POI l to be promoted, TPD aims at finding a set of influential users, who can generate more users to visit l in the future, and TCD is to find a set of potential users, who will visit l in the future. To deal with TPD and TCD, we propose a novel graph embedding method, LBSN2vec. The main idea is to jointly learn a low dimensional feature representation for each user and each location in an LBSN. Equipped with learned embedding vectors, we propose two similarity-based measures, Influential and Visiting scores, to find potential targeted propagators and customers. Experiments conducted on a large-scale Instagram LBSN dataset exhibit that LBSN2vec and its variant can significantly outperform well-known network embedding methods in both tasks.

Keywords: location-based social networks; promotion propagators; potential customer; location promotion; embedding learning

1. Introduction

In recent years, location check-ins at Points-of-Interests (POIs) on social media has become a living habit. People love sharing their life on Facebook or Instagram. Whether it is traveling or tasting food, they will check-in or share photos to announce the locations. Then location-based Social Networks (LBSN) connecting user social relationships and their check-ins at POIs can be generated. LBSN enables location-aware recommendation applications, such as POI recommendation [1–7] and potential customer recommendation [8,9].

Location promotion is one of the essential problems in LBSN [10,11], and the basic goal is to find the set of seeds (i.e., information *propagators*) to maximize the number of users (i.e., *customers*) to visit the target POI [10–14]. An effective location promotion algorithm can benefit the location service providers (e.g., restaurants and stores) to search the most influential promoters for targeted marketing. Another beneficial audience is users who are willing to be recruited as location promoters, i.e., they can earn money from the location service providers. In existing studies, however, both the selection of propagators and the estimation of customers rely on historical user-location interaction data (e.g., check-in records). The identified seeding propagators may not be influential in the future. In real practice, it is usually expected that the identified propagators can truly generate more users to visit the target POI (i.e., more customers) in the future, as discussed in existing studies [10,11,15]. In other words, existing location promotion lacks predictability when finding propagators and customers.

To consider the predictability in location promotion, this paper proposes to solve two novel problems. The first is *Targeted Propagator Discovery* (TPD) while the second is *Targeted Customer Discovery* (TCD). The TPD problem is that given a target location that intends to initiate a promotion campaign, the goal is to *predict* which users are influential with respect to the target location during the upcoming promotion period. Specifically, we aim to find which users in LBSN can generate more users to visit the target location when the promotion campaign starts. To the best of our knowledge, we are the first attempt to deal with the problem of TPD. TPD has some differences from location promotion. TPD aims to find the set of users as propagators who will actually generate more other users visiting the target POI *in the future*, but conventional location promotion is to find the set of users who have a large influence on the target POI using historical check-in data *in the past*. Users who are influential in historical check-in data might not imply they will still be influential in the future, especially on activating other users to visit the target POI. Since POI holders usually have limited funding, accurately finding future propagators can bring more revenue for the target POI. On the other hand, the TCD problem is to predict which users will visit the target location during the upcoming promotion period. With TCD, the promotion initiator of the given POI can perform targeted marketing to boost the possibility of a user visiting. In addition, with TCD, the POI holders can first understand the common traits of the potential customers in advance so that the marketing strategy can be well constructed.

For the problem of TPD, it concerns the scenario that the POI holders, who aim to market or to promote sales, wonder which users are not only influential but also can attract more people to visit the POI in advance. Then POI holders can hire more influential people to market the POI that can increase the visited people to create more business opportunities. For example, a restaurant holder only has limited funds to employ three bloggers to advertise the restaurant. If we can find the potential candidates who have more attraction ability in the social network, by performing marketing based on these candidates, the POI will bring greater benefits to the restaurant holder. In other words, every venue would like to promote itself by recruiting some influential propagators under a limited budget. Therefore, it is crucial to devise a prediction-based method that can accurately identify such propagators for effective location promotion.

Moreover, for the problem of TCD, it concerns that the POI holders, who aim to advertise their POIs, want to perceive who the potential customers are when designing the promotion plan. The POI holders can analyze the potential customers to know the characteristics they have, and develop an effective marketing strategy that can successfully attract such potential customers. For example, if we can find the potential users who have a larger chance of visiting the restaurant, then the restaurant can send some discount coupons or send some mails to advertise the POI so that their probabilities of visiting the POI can be remarkably boosted. That said, not all users will be interested in the targeted POI, and our goal is to develop a prediction-based method to find those potential customers. By doing so, the POI holders can avoid putting resources on irrelevant users, and concentrate more on potential customers.

We take advantage of social relations and user-POI interactions in LBSN to make predictions in both TPD and TCD problems. While existing studies estimate the number of activated users via information propagation [11,16,17], we think such simulation-based influence estimation cannot reflect the potential of a user becoming influential, i.e., affecting other users to visit the target POI. Therefore, we propose a network embedding method, LBSN2VEC, to depict users and their interactions with POIs. The idea is to jointly embed users in the social network as well as their check-in records at POIs into a low-dimensional space. The embedding space is expected to preserve user preferences as well as user-user interactions in visiting POIs. In the embedding space, two users close to each other can reflect that: (a) they have similar preferences of POI visiting, and (b) they have higher potential to affect one another. That is, one of them visiting a new POI tends to give higher motivation for the other also to visit that POI. In addition, since check-in records are also used in learning embeddings, the embedding vectors of users and POIs are comparable. And the distance between a user and a POI can imply user preference in visiting that POI. A closer distance between a user and a POI can reflect

that the user has a higher desire to visit the POI. Consequently, for a given POI, we propose some similarity measures to retrieve influential users as propagators and potential users as customers.

Here we would like to elaborate the underlying reasons that we adopt the embedding learning approach to solve the proposed tasks. To better motivate the proposed method, we create Table 1 to systematically compare the neural network embedding-based approach with the simple statistical scoring heuristics in different aspects. First, both our TPD and TCD tasks aim to predict the future behaviors of users with respect to the given targeted POI. Existing node scoring measurements, such as the weighted neighborhood size and the number of involved geodesics, mainly rely on a variety of statistical calculations, less regarding predictability. The proposed embedding learning methods can better model how users interact with POIs, and thus properly work for predicting future propagators and customers. Second, the statistical methods, such as the number of geodesics ending in targeted POI, multiplying the neighborhood size or centrality score, cannot model the semantic behaviors of users. That said, although two users could have totally different visited POIs, these POIs can belong to similar types. If we do not model the semantic behaviors of users via exploiting their high-order interactions in the check-in graph, the user influence and user's predictability to the target POI cannot be well learned. The proposed embedding learning method is to better depict the high-order semantic relationships between users and items in the vector space. This is the main reason that we adopt the embedding learning method. Third, in addition to representing users by their visited POIs, we think the POIs can also be depicted by those users who visit there. Rather than quantitatively considering POIs in a statistical manner, we jointly project users and locations into their common embedding space so that we can estimate and exploit their distances to better distinguish their affinity to the targeted POIs.

Table 1. Comparison between simple scoring heuristics and embedding learning-based approach.

	Simple Scoring Heuristics	Embedding Learning Approach
Example Methods	(a) Number of involved geodesics (b) Weighted neighborhood size (c) Centrality-enhanced measures	(a) The proposed LBSN2vec (b) Typical node embedding (e.g., node2vec [18]) (c) Transition embedding (e.g., POI2vec [14])
Predictability	No. Just simple statistics.	Yes. Make prediction based on learned embeddings.
Feature Representation	No. Do not characterize and distinguish users and POIs from one another	Yes. Learning the semantic distinguishable high-order interactions between users and POIs
Comparable	No. User and POIs are different information types.	Yes. Users and items are mapped into a common embedding, and thus comparable
Applicable Scenarios	(a) Finding popular POIs and users. (b) Mining frequent user-POI interactions.	(a) Targeted propagator discovery (TPD) (b) Targeted customer discovery (TCD) (c) POI recommendation systems

Below summarizes the contributions of this paper.

- We formulate two novel problems: Targeted Propagator Discovery (TPD) and Targeted Customer Discovery (TCD). To the best of our knowledge, we are the first study that proposes the problem of TPD.
- We propose a network embedding model, LBSN2VEC, to jointly learn the feature representations of POIs and users based on the social network and user-POI check-in data in LBSN. A more effective variant, PLBSN2VEC, is also presented to further boost the predictability of the embedding space.
- Experiments were conducted on a large-scale Instagram dataset to examine the performance of LBSN2vec and PLBSN2vec. The results exhibit that LBSN2vec can outperform well-known network embedding methods in both TPD and TCD tasks.

This paper is organized as below. We first present the problem formulation in Section 2. Then we describe the proposed LBSN2vec and PLBSN2vec in Section 3. Section 4 demonstrates the experimental setup and presents a discussion of the results. Section 5 concludes this work.

2. Related Work

The relevant studies can be divided into two categories, location promotion and network representation learning in graphs. We will first describe both, and discuss their comparisons with our work.

Location Promotion in LBSNs. Location promotion is to find a set of users as seeding promoters such that the number of influenced users visiting the targeted POI can be maximized. There are several typical methods in this research direction. Likhvani et al. [10] developed the model named LoCaTe to quantify the influence of a location and a user by modeling user-location interaction patterns, and categorical and temporal semantics between users and locations. Liao et al. [12] aim at modeling how users interact with each other via their friendships, and accordingly devise a user proximity score for recommending companions for an activity. Bouros et al. [13] incorporate the geographical region information into the identification of influential users, and develop a new LBSN influence propagation method to achieve the goal. POI2Vec [14] is designed to predict users who will visit a given location within a future period. POI sequential transition and user preference are used to learn the representation of POI so that potential visitors can be effectively predicted. Zhu et al. [11] also aim to select a set of seeding users and expect them to attract most other users to visit a given location. Its novelty is the modeling of user mobility from check-in records so that the location-aware propagation probabilities can be better derived.

The main difference between our work and existing studies in this category is two-fold. First, past methods on location promotion utilize historical check-in records to find the most influential users who can attract the most other users to visit a targeted location. The influence of users is estimated based on historical data. However, these methods cannot predict whether the identified influential users can truly visit the location in the future. In other words, these methods are not to predict the *future* influence of users for location promotion. Second, although the method proposed by Zhu et al. [11] can effectively predict whether a user will visit the targeted location in the future, it is not to identify the influential users who can better serve as seeding promoters, attracting more users to visit the targeted location. In short, our work is the first work that predicts future user influence that can bring more other users to visit the targeted location in the future. Note that the main difference between the proposed propagators and the so-called influencers in existing studies is that previous influencers are identified based on only the social network structure via diffusion simulation. The identified influencers cannot be ensured to have the same influence in the future. That said, the mining of influencers in the literature does not consider the predictability. We argue that for real applications, the predictability of whether influencers can truly activate more users should be taken into account. We treat the influencers with predictability as propagators to be detected in our work.

Network Representation Learning in Graphs. Network representation learning, also known as node embedding learning, aims to generate a low-dimensional vector for each node in a graph. The goal is to preserve the neighborhood proximity between nodes in the embedding vectors of nodes. Two nodes that share similar neighbors are supposed to have similar vectors in the embedding space. There are three typical studies in network representation learning, node2vec [18], LINE [19], and GraphGAN [20]. The node2vec proposes to sample the neighbors of a particular node via a hybrid BFS and DFS mechanism, and apply the Skip-gram model [21] to generate the embedding vectors. LINE aims to preserve both the local and global network structures, i.e., the first- and second-order pairwise proximities, for nodes in the embedding space. GraphGAN applies generative adversarial networks (GAN) for network representation learning. Given a node in a network, the generative model is to fit its connectivity distribution over all nodes, and produces fake embeddings to make the discriminative model wrong. The discriminative model aims to determine whether the generated node embedding is true or artificial, i.e., generated by the generative model. In our experiments, we will compare the proposed methods with these three models of network presentation learning.

The differences between the proposed LBSN2vec and existing graph embedding methods consist of three parts. First, our LBSN2vec is able to effectively exploit two graphs, i.e., the social network and the check-in network, to generate the node embeddings by jointly considering social neighborhood and check-in neighborhood. However, the existing method takes advantage of only the social network. Being able to incorporate multi-model information in a united framework is the key to LBSN2vec. Second, in LBSN2vec, we introduce a penalization mechanism to avoid the random walk traversing the same node so that nodes with higher distinguishability can be sampled. Such a penalization cannot be achieved by existing methods. Third, in the experiments, we have found that when having our extensive design from existing methods, LBSN2vec and its variant PLBSN2vec can improve the performance on both tasks of TPD and TCD.

Recent Advances in LBSN. Recent advances in mining location-based social networks focus on recommendation point-of-interests for users. Das et al. [22] classify immediate and distant social relationships based on check-in data for better POI recommendation. A hierarchical attention mechanism (HAM) [23], which effectively selects the more representative LBSN data instances for training, is devised to perform efficient personalized POI recommendation. DSPR [24] further leverages user preferences and real-time demand, along with an attention-based recurrent neural network, for POI recommendation. While labeled data are scarce in practice, a semi-supervised learning framework HisRect [25] is developed to take advantage of unlabeled data to recommend the next POIs. SACRA is a self-attentive prospective customer recommendation framework [26] that makes comparisons between users' check-ins by adversarial training so that POI recommendation can be more robust. Li et al. [27] further combine metric learning and few-shot learning to fully utilize user check-ins and better learn the semantic matching between users and POIs. GGLR [28] is a graph-based geographical latent representation model to capture non-linear geographical influences based on the user-POI check-in graph.

Although these are successful proposals for point-of-interest recommendation for users, none of them discuss how to apply advanced neural network and representation learning methods for discovering influential propagators and predicting potential customers. We believe existing methods can accurately predict the next POIs. Since our problem is significantly different from POI recommendation, i.e., estimating the future influence of users regarding a targeted location in our work, a novelty of our work can be the first to adopt representation learning for predicting propagators and customers in location-based social networks.

3. Problem Formulation

In this section, we formulate the problem of Targeted Propagator Discovery and Targeted Customer Discovery. The Targeted Propagator Discovery aims to find the set of users who will generate more people to visit the target POI in the future. Moreover, the aim of Targeted Customer Discovery is to find the set of users who will visit the target POI in the future.

We first present the notations. Let $G_1 = (U, E_1)$ be a social network, where U is the set of users and $E_1 = \{e_{ij} = (u_i, u_j)\}$, $u_i, u_j \in U$. Also let (u, p, t) represent a check-in record, which means user u visits POI p at the time t . We denote a check-in network as $G_2 = (V, E_2)$, where $V = U \cup P$, P is the set of POIs and $E_2 = \{e_{up} = (u, p)\}$, where $u \in U$ and $p \in P$.

Problem 1 (Targeted Propagator Discovery, TPD): Given an LBSN containing social network G_1 and check-in network G_2 , a target POI p and the current time t_c , the goal of TPD is to find the k users who will generate more people to visit POI p in the time period $[t_c, \infty]$.

Problem 2 (Targeted Customer Discovery, TCD): Given an LBSN with social network G_1 and check-in network G_2 , a target POI p and the current time t_c , the goal of TCD aims to find a set of users who will visit POI p in the time period $[t_c, \infty]$.

4. The Proposed Method

The proposed method consists of two parts. The first is to learn the embedding vectors of users and POIs in Section 3, which describe how we generate “contexts” of nodes in the social network and the check-in network, and how we learn the embeddings, respectively. That said, we will learn a mapping function $f : V \rightarrow \mathbb{R}^d$ from users and POI nodes in G_1 and G_2 and output the d -dimensional embedding vectors. The second is a scoring function to estimate the similarity between users and POIs so that both TPD and TCD can be solved in Section 3.

The main properties preserved by LBSN2vec are two parts. First, the similarity between users with regard to their check-in POIs is preserved by LBSN2vec. Two users possessing more visited check-in POIs tend to be close to each other in the embedding space generated by LBSN2vec. Such a property cannot be well preserved by existing graph embedding methods. Second, since LBSN2vec jointly considers neighboring users and visited POIs to be the context of a node, the mix of two diverse contexts can better deal with the data sparsity of either part. That said, our LBSN2vec can preserve or alleviate the sparsity property in location-based social network data.

4.1. Random Walk Mechanism

We take advantage of the skip-gram model [21] to learn the feature representation of users and POIs. Let $N_R(u) \subset U$ be the set of user u 's neighbors in graph G_1 . To have the contexts in the skip-gram model, we generate $N_R(u)$ through a devised random walk mechanism. Below we first present the probability associated with each edge, then describe the random walk mechanism.

We consider that nodes with strong social connections with each other tend to be treated as neighbors. Thus we introduce two *affinity* scores to quantify the social strength between users i and j . The first is *social affinity* $\pi_s(i, j)$, defined as:

$$\pi_s(i, j) = \frac{|N_{G_1}(i) \cap N_{G_1}(j)|}{|N_{G_1}(i) \cup N_{G_1}(j)|},$$

where G_1 is the social network, $N_{G_1}(i)$ is the neighborhood of user i in G_1 . The second is *co-visiting affinity* $\pi_c(i, j)$, defined as:

$$\pi_c(i, j) = \frac{|N_{G_2}(i) \cap N_{G_2}(j)|}{|N_{G_2}(i) \cup N_{G_2}(j)|},$$

where G_2 is the check-in network, $N_{G_2}(i)$ is the neighborhood of user i in G_2 .

Since π_c shows the affinity based on user-POI interactions, if there are more co-visited POIs between users, π_c gets higher. Two users are more likely to possess similar POI visiting preferences in the future. On the other hand, π_s shows the affinity based on the social network. If two users have more common friends, π_s gets higher, i.e., the two users are more likely to affect each other. Then we define the edge weight (probability) w_{ij} between users i and j in the social network by combining social affinity and co-visiting affinity, given by:

$$w_{ij} = \pi_s(i, j) + \pi_c(i, j), \forall e_{ij} \in E_1,$$

where E_1 is the edge set of social network G_1 .

Here we present examples to demonstrate the derived neighboring POIs sampled by random walks of co-visiting affinity and social affinity, in Figure 1, in which different colors mean various venue categories. We can find that the random walk by co-visiting affinity tends to collect neighbors with the same category as the target POI. The random walk by social affinity has a higher potential to bring neighbors with diverse venue categories. Hence, we would like to have both effects so that the representation learning can generate similar embeddings that not only encode similar categories but also distinguish from each other due to different social interests.

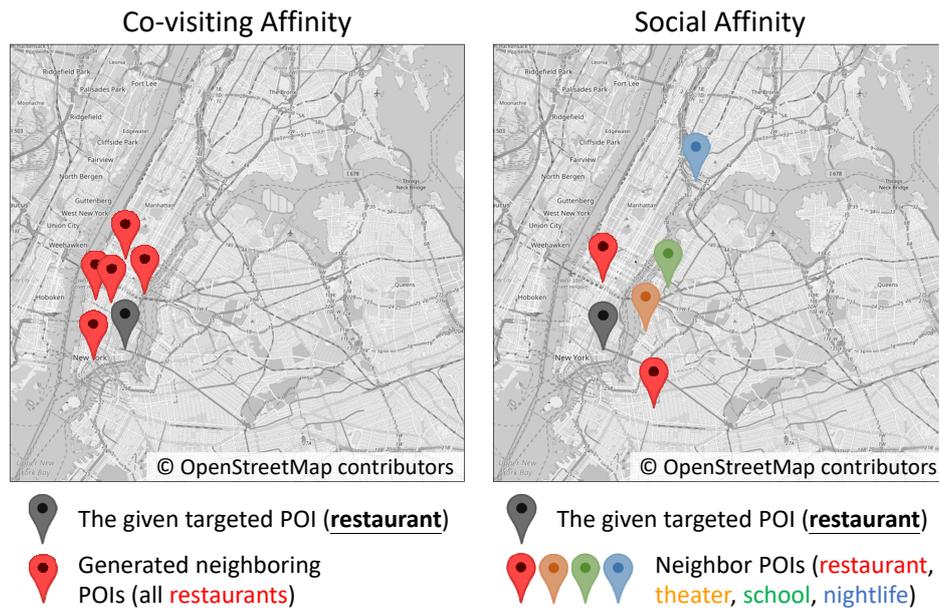


Figure 1. The given sample targeted point-of-interest (POI) and their POI neighbors, along with POI categories by various colors, based on random walk sampling of co-visiting affinity (left) and social affinity (right). Note that for these two pictures, their base maps and data are generated from [OpenStreetMap](#) and [OpenStreetMap Foundation](#).

We also propose a penalization-based LBSN2vec (PLBSN2vec) by adding a penalty parameter in the random walk: a return parameter ρ ($0 < \rho < 1$) and a parameter $|Path_{nr}|$ effect. The return parameter ρ is the penalty value to avoid traversing the same nodes. In addition, we think that near-by nodes of a user tend to have similar preferences. Hence, we let those close to each other have a high possibility of being traversed as the neighbors. Given the starting user u , the next user n , and the previous user r , we devise the penalty weight p_{nu} as:

$$p_{nu} = \begin{cases} \frac{1}{|Path_{nr}|} \cdot w_{nu} & \text{if } |Path_{nr}| \geq 1 \\ \rho \cdot w_{nu} & \text{if } |Path_{nr}| = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $|Path_{nr}|$ is the shortest path length between n and r . Lower values of return parameter ρ make the walk have a lower opportunity to repeat the previous user. By contrast, a larger return parameter ρ will increase the chance of walking back to the previously traversed users. In addition, longer paths between nodes get a larger penalty in $|Path_{nr}|$, which will decrease the probability.

Then we present the random walk mechanism. Formally, given the current node c , we simulate a random walk of length l . Let r_i be the i -th node in the random walk. For LBSN2vec, we define the conditional probability of the next node n given the current node c :

$$P(r_i = n, r_{i-1} = c) = \begin{cases} \frac{w_{nc}}{Z} & \text{if } e_{nc} \in E_1 \text{ (for LBSN2vec)} \\ \frac{p_{nc}}{Z} & \text{if } e_{nc} \in E_1 \text{ (for PLBSN2vec)} \\ 0 & \text{otherwise} \end{cases}$$

where Z is the normalizing constant. After simulating the random walk with length l and window size w , we can obtain the neighbors of a node, which are considered as the so-called “contexts” in the skip-gram model.

4.2. Embedding Learning

This section presents the embedding learning of LBSN2vec and PLBSN2vec based on the skip-gram model architecture. We adopt a three-layer neural network [21,29], consisting of an input layer, a hidden layer, and an output layer. We consider the matrix f from the input layer to the hidden layer to be the projection matrix, and there is a difference matrix (denoted as f') between the hidden and output layers. Let $f : v \rightarrow \mathbb{R}^d$ be the mapping function from node $v \in V$ to its embedding vector, where f is the projection matrix of size $|V| \times d$, and d is the dimensionality of embeddings ($d \ll |V|$). The output layer (denoted by y) is the vector of size $|V| \times 1$. We apply the softmax function to define the conditional probability of user u_i given the input user u_l as:

$$y_{u_i,j} = p(u_{u_i,j}|u_l) = \frac{\exp(f(u_l) \cdot f'(u_j))}{\sum_{v \in V_2} \exp(f(u_l) \cdot f'(v))}$$

where $u_i \in N(u_l)$. Skip-gram is the model that predicts the context based on the given word (node). We replace the context with the neighbors generated through the random walk mechanism in Section 3. The training objective O is to maximize the conditional probability of neighborhood given the user, given by:

$$O = \max_{u \in U} \sum p(N(u)|u) \equiv \max_{u \in U} \sum \log(p(N(u)|u))$$

where $N(u)$ is the neighborhood of user u . $N(u)$ can be generated based on either social network G_1 or check-in network G_2 . We optimize the objective function by stochastic gradient descent (SGD).

4.3. Making Prediction via Similarity

For the TPD task, we aim to find a set of users who will generate more people to visit the given POI. We consider quantifying two factors to determine whether a user will visit the given POI. One is user preference $uf(u_1, u_2)$, given by:

$$uf(u_1, u_2) = sim(f(u_1), f(u_2)),$$

where $u_1, u_2 \in U$. The other is POI preference $pf(u, p)$, given by:

$$pf(u, p) = sim(f(u), f(p)),$$

where $u \in U, p \in P, f(\cdot)$ is the embedding vector, and $sim(\cdot)$ is the cosine similarity to compute the relation between two input vectors. Two users with similar preferences will have similar embedding vectors. So user preference is devised to depict the tendency of being influential between users. Moreover, if the similarity between the embedding vectors of a user and a POI is high, we consider that the user has a higher chance of visiting the POI. Hence, POI preference is designed to estimate the POI visiting potential/desire of a user.

To find the future targeted propagators, we develop a *contagion* scoring measure by combining user preference and POI preference.

$$contagion(u) = \sum_{u_i \in N_{G_1}(u)} uf(u, u_i) \cdot pf(u_i, p).$$

For the TPD task, we rank all users' contagion scores, and select the top- K users as the discovered propagators, i.e., those that can generate the most people to visit the given POI.

For the TCD task, we devise a *visiting* scoring measure considering POI preference:

$$visiting(u) = pf(u_i, p).$$

Then we rank all users' scores, and select the top- K users as the potential customers for the given POI.

The novelty of our similarity-based measures lies in its inductive predictability on LBSN. First, measuring the similarity to identify targeted propagators and potential customers is an unsupervised learning approach. Since we do not have any label data about future propagators and customers, the similarity-based unsupervised method can properly generate the results without supervision information. This would be especially useful for location promotion because we cannot have the labeled promotion data before campaigns.

5. Experimental Evaluation

We conduct experiments to examine whether the proposed LBSN2vec and PLBSN2vec can better model user preferences and user-POI interactions in TPD and TCD tasks, comparing with well-known network embedding methods. Collected Instagram data are used for the experiments (the dataset can be crawled via Instagram API: <https://www.instagram.com/developer/>—due to the privacy issue established by Instagram, we cannot provide the data for download). The dataset contains check-in records and user–user followships collected in New York City from Jan 2012 to Nov 2015. There are 994,412 user-POI check-ins and 1,048,575 followship relations. We randomly select 200 POIs and repeat the random selection up to 50 times for the experiments. We use the first 80% chronological check-ins for training and the remaining 20% for testing. We compare LBSN2vec and PLBSN2vec with several well-known network embedding methods: node2vec [18], LINE [19], and GraphGAN [20]. They all map nodes in a network to embedding vectors. The embedding vectors derived from all these methods are fed into the proposed scoring function for both TPD and TCD tasks. We should notice that the difference between our LBSN2vec and PLBSN2vec lies in that the former considers only w_{ij} as the random walk probability while the latter further imposes the penalized parameters.

We choose three common measures in the information retrieval field as the evaluation metrics [30]: (a) Precision@ N , (b) Recall@ N , and (c) NDCG@ N :

$$\begin{aligned} \text{Precision@}N &= \frac{\#\{Pred(N) \cap GT(N)\}}{N} \\ \text{Recall@}N &= \frac{\#\{Pred(N) \cap GT(200)\}}{N} \\ \text{NDCG@}N &= \frac{DCG@N}{IDCG@N} \end{aligned} \quad (1)$$

where $Pred(N)$ is the top- N predicted ranking, and $GT(N)$ top- N ground-truth ranking. Furthermore, $DCG@N = \sum_{i=1}^N \frac{rel_i}{\log_2(i+1)}$ and $IDCG@N = \sum_{i=1}^N \frac{N+1-i}{\log_2(i+1)}$. Precision aims at examining whether a method can accurately rank the ground truth at the top position while Recall is used to quantify whether all of the ground truth can be retrieved in the ranking list. The measures of precision and recall are trade-offs between each other. They must be seen together to understand the effectiveness of a model. As for NDCG, it is to estimate how the predicted ranking is close to the ground-truth ranking, which cannot be captured by Precision and Recall.

For all methods, we use the same default parameters listed below if they have the settings: the embedding dimension $d = 100$, the number of walks per node = 10, the walk length $l = 80$, and the neighborhood size $w = 6$. In our PLBSN2vec methods, we set the return parameter $\rho = 0.8$ and chose the user-POI weighting parameter $\alpha = 0.1$ by default.

Results of TPD. We present the results of Precision@ N ($N = 5, 10, \dots, 50$) in Figure 2. The results of Precision@10, Recall@10, and NDCG@10 are also shown in Table 2. It can be apparently found that LBSN2vec and PLBSN2vec outperform the other embedding methods by varying N . In addition, as N gets increased, PLBSN2vec is able to stably have the higher Precision values. Among the competing methods, we find that GraphGAN has the worst performance, as it cannot capture the contagion between users and POIs. We think our methods lead to better performance because two key factors are

considered. One is the modeling of user–user contagion, and the other is user-POI visiting preference. Such two factors can be better modeled into the embedding vectors of our proposed LBSN2vec and PLBSN2vec.

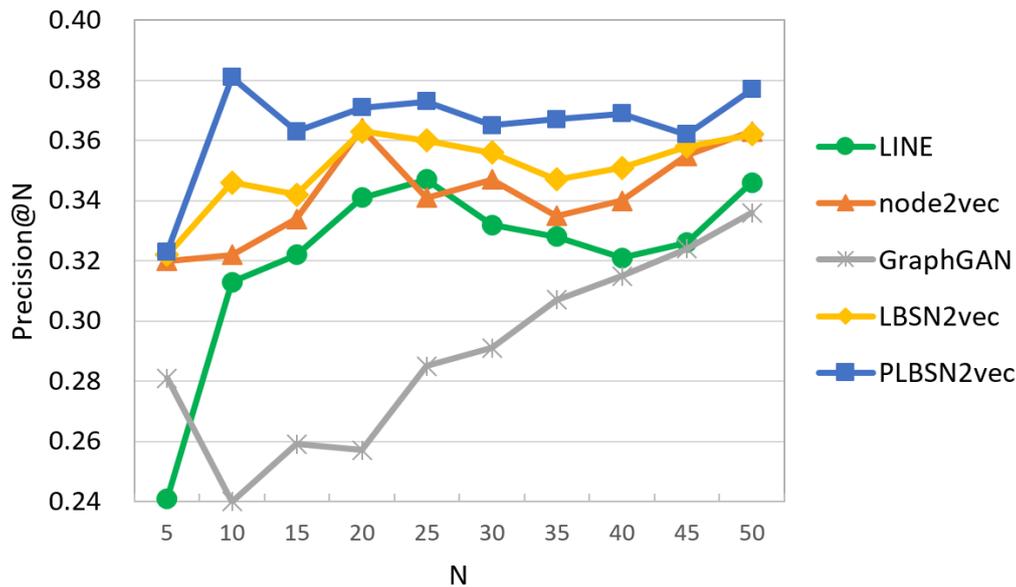


Figure 2. Results of Precision@N for the Targeted Propagator Discovery (TPD) task.

Table 2. Results in P@10, R@10, and NDCG@10 for the TPD task.

	P@10	R@10	NDCG@10
Node2vec	0.320	0.185	0.741
LINE	0.310	0.195	0.610
GraphGAN	0.240	0.169	0.675
LBSN2vec	0.350	0.197	0.760
PLBSN2vec	0.380	0.216	0.753

Results of TCD. In TCD, we believe the promotion initiator would care more about whether all potential customers can be found. Hence, we report Recall@N in this part. Figure 3 shows the results in Recall@N by varying N. We can see the Recall values with different embedding methods. Our LBSN2vec and PLBSN2vec are able to significantly outperform the other network embedding methods. We think the superiority of our methods comes from the derived embedding space that better utilizes the information between user–user interactions and user-POI visiting records in location-based social networks, which is crucial in modeling the potential of future POI visiting.

Limitations. Here we would like to discuss the limitations of the proposed methods for effectively solving TPD and TCD tasks. First, the proposed embedding learning methods, LBSN2vec and PLBSN2vec, are based on historical check-in records, i.e., existing interactions between users and POIs. If users tend to visit some particular POIs, it would be possible that the recommended propagators and potential customers, with respect to one of these POIs, can to some extent be similar. In other words, the “filter bubble” effect could happen. We would leave the solution to filter bubble as future work because there is a critical challenge in striking a balance between having accurate predictions and fair recommendations. Second, the embedding learning-based approach via a neural network requires a rich set of user-POI interactions to generate effective representations. While it is common to have some popular POI types and popular POIs, it would be natural to have better performance for popular venues and types. Third, different POIs have a variety of behaviors of users. Some are frequently visited by similar groups of users, and some are visited by various types of users. No matter which types of POIs are given, our proposed method can recommend proper propagators and customers.

Due to the requirement of rich data for better prediction performance, we think our method may be less suitable for venues visited by users with diverse behaviors.

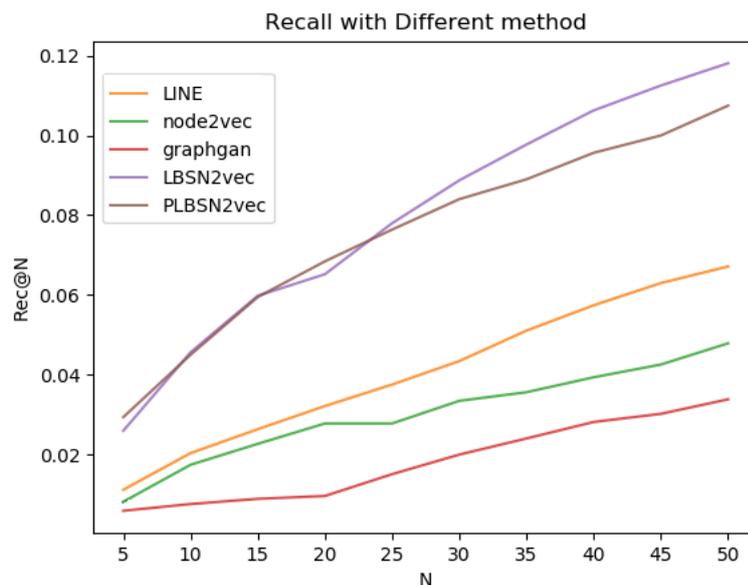


Figure 3. Results in Recall@N for the TCD task.

6. Conclusions and Discussion

In this paper, we propose two novel problems, Targeted Propagator Discovery (TPD) and Targeted Customer Discovery (TCD), in the context of location promotion. To the best of knowledge, we are the first to find future potential propagators in location-based social networks. We propose two network embedding methods, LBSN2vec and PLBSN2vec, along with a modified random walk mechanism based on the Skip-gram model. We also devise two scoring functions with the estimation of user preference and POI preference for TPD and TCD tasks. Experiments conducted on a large-scale Instagram dataset exhibit superior performance, compared to several well-known network embedding methods.

While effectively solving both TPD and TCD tasks by the proposed LBSN2vec and PLBSN2vec, two implications should be discussed. First, one may concern that the spatio-temporal footprints of users can be obtained and revealed by the POI holders who are performing the algorithms, leading to the issue of privacy leakage. We should highlight that the POI holders can only access the discovered propagators and potential customers. The original data and the footprints of general users are still maintained by the service providers who have an obligation to keep the prediction results confidential. Therefore, every POI holder cannot trace users. Second, one may also think that the POI holder can directly contact the recommended targeted propagators. Such an action could happen indeed. Here our algorithms leave the mechanism design on how to contact propagators and potential customers to the service providers. Whether and how the users react to being propagators or receiving the discount coupons should also be properly devised.

Author Contributions: Conceptualization, Y.-C.C. and C.-T.L.; methodology, Y.-C.C. and C.-T.L.; validation, Y.-C.C. and C.-T.L.; formal analysis, Y.-C.C. and C.-T.L.; investigation, Y.-C.C. and C.-T.L.; writing—original draft preparation, Y.-C.C. and C.-T.L.; writing—review and editing, Y.-C.C. and C.-T.L.; supervision, C.-T.L.; funding acquisition, C.-T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Ministry of Science and Technology (MOST) of Taiwan under grants 109-2636-E-006-017 (MOST Young Scholar Fellowship) and 109-2221-E-006-173, and also by Academia Sinica under grant AS-TP-107-M05.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Feng, S.; Li, X.; Zeng, Y.; Cong, G.; Chee, Y.M.; Yuan, Q. Personalized Ranking Metric Embedding for Next New POI Recommendation. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI '15, Buenos Aires, Argentina, 25–31 July 2015; pp. 2069–2075.
2. Li, X.; Cong, G.; Li, X.; Pham, T.A.N.; Krishnaswamy, S. Rank-GeoFM: A Ranking based Geographical Factorization Method for Point of Interest Recommendation. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15, Santiago, Chile, 9–13 August 2015; pp. 433–442.
3. Lian, D.; Zhao, C.; Xie, X.; Sun, G.; Chen, E.; Rui, Y. GeoMF: Joint Geographical Modeling and Matrix Factorization for Point-of-interest Recommendation. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA, 24–27 August 2014; pp. 831–840.
4. Zhang, J.D.; Chow, C.Y. GeoSoCa: Exploiting Geographical, Social and Categorical Correlations for Point-of-Interest Recommendations. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15, Santiago, Chile, 9–13 August 2015; pp. 443–452.
5. Zhang, W.; Wang, J. A Collective Bayesian Poisson Factorization Model for Cold-start Local Event Recommendation. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, Sydney, Australia, 10–13 August 2015; pp. 1455–1464.
6. Li, H.; Ge, Y.; Hong, R.; Zhu, H. Point-of-Interest Recommendations: Learning Potential Check-ins from Friends. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, San Francisco, CA, USA, 13–17 August 2016; pp. 975–984.
7. Xie, M.; Yin, H.; Wang, H.; Xu, F.; Chen, W.; Wang, S. Learning Graph-based POI Embedding for Location-based Recommendation. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, Indianapolis, IN, USA, 24–28 October 2016; pp. 15–24.
8. Liu, Y.; Zhao, P.; Sheng, V.S.; Li, Z.; Liu, A.; Wu, J.; Cui, Z. RPCV: Recommend Potential Customers to Vendors in Location-Based Social Network. In Proceedings of the International Conference on Web-Age Information Management, WAIM '15, Qingdao, China, 8–10 June 2015; pp. 272–284.
9. Sarwat, M.; Eldawy, A.; Mokbel, M.F.; Riedl, J. PLUTUS: Leveraging Location-Based Social Networks to Recommend Potential Customers to Venues. In Proceedings of the IEEE 14th International Conference on Mobile Data Management, MDM '13, Milan, Italy, 3–6 June 2013; pp. 26–35.
10. Likhvani, A.; Bedathur, S.; Deepak, P. LoCaTe: Influence Quantification for Location Promotion in Location-based Social Networks. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI '17, Melbourne, Australia, 19–25 August 2017; pp. 2259–2265.
11. Zhu, W.Y.; Peng, W.C.; Chen, L.J.; Zheng, K.; Zhou, X. Modeling User Mobility for Location Promotion in Location-based Social Networks. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, Sydney, Australia, 10–13 August 2015; pp. 1573–1582.
12. Liao, Y.; Lam, W.; Jameel, S.; Schockaert, S.; Xie, X. Who Wants to Join Me?: Companion Recommendation in Location Based Social Networks. In Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval, ICTIR '16, Newark, DE, USA, 12–6 September 2016; pp. 271–280.
13. Bouros, P.; Sacharidis, D.; Bikakis, N. Regionally Influential Users in Location-aware Social Networks. In Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '14, Dallas, TX, USA, 4–7 November 2014; pp. 501–504.
14. Feng, S.; Cong, G.; An, B.; Chee, Y.M. POI2Vec: Geographical Latent Representation for Predicting Future Visitors. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI '17, San Francisco, CA, USA, 4–9 February 2017; pp. 102–108.
15. Wu, X.; Fu, L.; Yao, Y.; Fu, X.; Wang, X.; Chen, G. GLP: A Novel Framework for Group-Level Location Promotion in Geo-Social Networks. *IEEE/ACM Trans. Netw.* **2018**, *26*, 2870–2883. [[CrossRef](#)]
16. Kempe, D.; Kleinberg, J.; Tardos, E. Maximizing the Spread of Influence Through a Social Network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, Washington, DC, USA, 24–27 August 2003; pp. 137–146.

17. Li, G.; Chen, S.; Feng, J.; Tan, K.L.; Li, W.S. Efficient Location-aware Influence Maximization. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14, Snowbird, UT, USA, 22–27 June 2014; pp. 87–98.
18. Grover, A.; Leskovec, J. Node2Vec: Scalable Feature Learning for Networks. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
19. Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; Mei, Q. LINE: Large-scale Information Network Embedding. In Proceedings of the 24th International Conference on World Wide Web, WWW '15, Florence, Italy, 18–22 May 2015; pp. 1067–1077.
20. Wang, H.; Wang, J.; Wang, J.; Zhao, M.; Zhang, W.; Zhang, F.; Xie, X.; Guo, M. GraphGAN: Graph Representation Learning with Generative Adversarial Nets. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI), AAAI '18, New Orleans, LA, USA, 2–7 February 2018.
21. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
22. Das, J.; Agarwal, A.; Majumder, S.; Mali, K. Location Based Scalable Recommendations Using Social Relationships. In *Communication Software and Networks*; Springer: Singapore, 2020; pp. 477–485.
23. Pang, G.; Wang, X.; Hao, F.; Wang, L.; Wang, X. Efficient point-of-interest recommendation with hierarchical attention mechanism. *Appl. Soft Comput.* **2020**, *96*, 106536. [[CrossRef](#)]
24. Wang, H.; Li, P.; Liu, Y.; Shao, J. Towards real-time demand-aware sequential POI recommendation. *Inf. Sci.* **2020**, *547*, 482–497. [[CrossRef](#)]
25. Li, P.; Lu, H.; Zheng, Q.; Li, S.; Pan, G. HisRect: Features from Historical Visits and Recent Tweet for Co-Location Judgement. *IEEE Trans. Knowl. Data Eng.* **2019**. [[CrossRef](#)]
26. Li, R.; Wu, X.; Wang, W. Adversarial Learning to Compare: Self-Attentive Prospective Customer Recommendation in Location Based Social Networks. In Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20, Houston, TX, USA, 3–7 February 2020; pp. 349–357.
27. Li, R.; Wu, X.; Wu, X.; Wang, W. Few-Shot Learning for New User Recommendation in Location-Based Social Networks. In Proceedings of The Web Conference 2020, WWW '20, Taipei, Taiwan, 20–24 April 2020; pp. 2472–2478.
28. Chang, B.; Jang, G.; Kim, S.; Kang, J. Learning Graph-Based Geographical Latent Representation for Point-of-Interest Recommendation. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, Bratislava, Slovakia, 15–18 September 2020; pp. 135–144.
29. Rong, X. word2vec Parameter Learning Explained. *arXiv* **2014**, arXiv:1411.2738.
30. Baeza-Yates, R.A.; Ribeiro-Neto, B. *Modern Information Retrieval*; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1999.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).