

# Learning from Your Network of Friends: a Trajectory Representation Learning Model based on Online Social Ties

Basma Alharbi and Xiangliang Zhang

King Abdullah University of Science and Technology (KAUST), KSA

basma.harbi@kaust.edu.sa; xiangliang.zhang@kaust.edu.sa

**Abstract**—Location-Based Social Networks (LBSNs) capture individuals whereabouts for a large portion of the population. To utilize this data for user (location)-similarity based tasks, one must map the raw data into a low-dimensional uniform feature space. However, due to the nature of LBSNs, many users have sparse and incomplete check-ins. In this work, we propose to overcome this issue by leveraging the network of friends, when learning the new feature space. We first analyze the impact of friends on individuals’s mobility, and show that individuals trajectories are correlated with those of their friends and friends of friends (2-hop friends) in an online setting. Based on our observation, we propose a mixed-membership model that infers global mobility patterns from users’ check-ins and their network of friends, without impairing the model’s complexity. Our proposed model infers global patterns and learns new representations for both users and locations simultaneously. We evaluate the inferred patterns and compare the quality of the new user representation against baseline methods on a social link prediction problem.

## I. INTRODUCTION

Due to the popularity of Location-Based Social Networks (LBSNs), data capturing *human mobility* and *social ties* became available at an unprecedented scale. In order to utilize such data and facilitate user (and location) similarity-based tasks, one must represent users/locations by a *uniform* low-dimensional feature space. Existing works employ vector space models to obtain a uniform representation, then reduce the high dimensionality using various techniques, [1], [2].

Incomplete and sparse trajectories cause a critical challenge during representation learning. A common practice in related work is to either eliminate short trajectories [3] or estimate missing locations [1]. However, elimination of short trajectories will remove a large portion of the users, given the power law distribution of users check-ins [3], and estimation of missing locations requires non-trivial work.

Different from existing studies, our work addresses this key challenge by proposing a data-fusion Probabilistic Graphical Model (PGM) that utilizes online social links while learning from incomplete trajectories. In essence, our proposed model, Social Trajectory Amplification and Representation learning model (STAR), infers latent mobility patterns from two data modalities: individuals’ trajectories (intrinsic factors) and online social links (extrinsic factors). STAR assumes that each location (e.g., a plaza) has a

distribution over *activities* (e.g., sunset concert, holiday celebrations), where in turn each *activity* has an attractiveness distribution over users. Activities denote latent patterns learned from users (and their network of friends) co-visiting the same locations, where no restriction is placed on temporal co-location of users. We evaluate the quality of user representation using a link prediction task.

**Overview of Contributions.** We develop a novel model, STAR, for mobility representation learning with the following highlights.

- 1) STAR considers both *extrinsic* (social network) and *intrinsic* (user trajectory) factors in learning mobility patterns, while previous work inferred mobility patterns considering only *intrinsic* motivations.
- 2) STAR provides new representations of locations and users in a space of latent *activities*, which facilitates the analysis of locations and users.
- 3) STAR has a comparable complexity to Latent Dirichlet Allocation (LDA) [4], where only a constant factor is added during the sampling process.
- 4) STAR is directly applicable to anonymized locations (IDs of locations), which are simple to prepare (no need for preprocessing and semantic labeling), and do not invade users’ privacy.

## II. RELATED WORK

Existing work on human mobility representation differs on how they treated locations and incomplete trajectories. [5] learns from semantically labeled locations of complete trajectories, while [1], [2] complete trajectories with estimates. Models in [6], [7] learn from incomplete semantically labeled trajectories with no attempt to compensate for missing data. The model in [3] learns from long enough trajectories of anonymized locations, without any treatment of incompleteness but just eliminating short trajectories.

To the best of our knowledge, our work is the first to introduce a model that learns from incomplete trajectories with anonymized locations by leveraging online social links. Note that incomplete trajectories are not necessarily short, while short trajectories are usually incomplete. Comparing to previous work, STAR considers extrinsic factors (social ties) while learning mobility patterns, as opposed to existing work which only focuses on individuals’ intrinsically motivated

mobility patterns. In this regard, our work introduces an important dimension to the human mobility framework. In addition, learning from anonymized locations minimizes effort and error involved in semantic labeling.

Additionally, our proposed model is closely related to two topic models, LDA with optimized parameters [8] and coherent LDA [9]. Our model utilizes the concept of optimized asymmetric location-activity prior similarly to [8] as well as incorporates the generalized Pólya Urn model in the learning process, similarly to [9]. Together, this combination of model elements serves to address the difficulties imposed by the special data attributes, discussed next, and the requirement of learning with social ties.

### III. MOBILITY AND SOCIAL TIES

#### A. Notations and Definitions

**Notation 1 (Bag-of-Users)** Let  $v \in \mathcal{V}$  be a user,  $c \in \mathcal{C}$  be a location, and  $\mathbf{X} \in \mathbb{N}^{|\mathcal{C}| \times |\mathcal{V}|}$  be a sparse matrix of location-user counts, where  $\mathbb{N}$  is the set of non-negative natural numbers and  $|\cdot|$  is the cardinality of the set. Let  $\mathbf{x}_c$  denotes the raw vector of all users' check-ins for location  $c$  and let  $\mathbf{x}_v^T$  denotes the column vector of all locations' check-ins for user  $v$ . Let  $x_{c,v}$  be the number of times user  $v$  visited location  $c$ .

**Notation 2 (User Adjacency Matrix)** Let  $\mathbf{A}^\kappa$  be a symmetric user-by-user matrix encoding users social links, where  $a_{v,v'}^\kappa = 1$  indicates that users  $v$  and  $v'$  are connected by a social tie. Superscript  $\kappa \in \mathbb{N}$  indicates the maximum degree of connection recorded in the matrix. For example, when  $\kappa = 1$ , i.e.,  $\mathbf{A}^1$ , a link between  $v$  and  $v'$  indicates that they are connected by a 1-hop degree (direct connection exists). For simplicity, we use  $\mathbf{A}$  instead of  $\mathbf{A}^1$  when  $\kappa = 1$ . When  $\kappa = 2$ ,  $a_{v,v'}^\kappa = 1$  means that they have a direct link or that they are connected via a 2-hop connection (friends of friends). An alternative explanation is that  $\mathbf{A}^\kappa$  encodes connected users with shortest path up to  $\kappa$  hops. We use  $\mathbf{a}_v^\kappa$  to denote the vector of user  $v$ 's  $\kappa$ -hop social links.

**Problem Definition. (Representation Learning)** Given trajectories represented as *bag-of-users*,  $\mathbf{X}$ , and online social ties,  $\mathbf{A}^\kappa$ , the goal is to learn a compact user and location representations. That is, given  $\mathbf{X}$  and  $\mathbf{A}^\kappa$ , learn  $\Theta \in \mathbb{R}_+^{|\mathcal{C}| \times K}$  and  $\Phi^T \in \mathbb{R}_+^{|\mathcal{V}| \times K}$ , for locations and users, respectively.  $K$  is the dimension of the new representation space such that  $K \ll \min(|\mathcal{V}|, |\mathcal{C}|)$ , and  $\mathbb{R}_+$  denotes the set of non-negative real numbers<sup>1</sup>.

#### B. Datasets and Preliminary Analysis

We study large-scale human trajectories extracted from LBSNs<sup>2</sup>. Precisely, we study two LBSNs, Gowalla (GW)

<sup>1</sup>Representation is restricted to non-negative for interpretability.

<sup>2</sup>This, however, should not limit our proposed model to data from LBSNs only as large-scale human trajectories from different sources exhibit similar attributes (such as CDRs), and thus we can leverage social ties to amplify incomplete trajectories in other datasources as well.

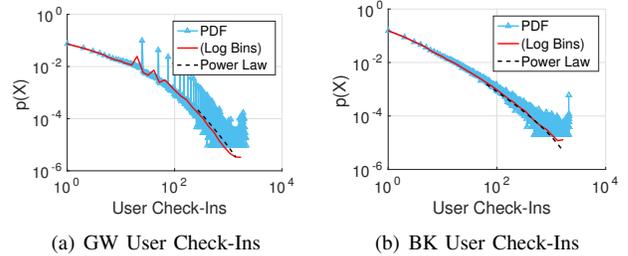


Figure 1. User and Location Check-In Distributions.

and Brightkite (BK), both publicly available [10]. Table I contains basic statistics of the two datasets. User trajectories are represented as a bag-of-users,  $\mathbf{X}$ , and undirected social links are represented as the user adjacency matrix,  $\mathbf{A}$ .

Table I BASIC DATA STATISTICS

		Gowalla	BrightKite
No. of Active Users	$ \mathcal{V} $	107,092	50,687
No. of Social Links	$ \mathbf{A}_{n,nz} $	456,830	194,091
No. of Locations	$ \mathcal{C} $	1,280,969	772,965
No. of Check-ins	$\sum_{c,v} \mathbf{X}$	3,981,334	1,072,965

We analyze the user check-in degree distribution, which measures the *total number of check-ins per user* and is shown for both datasets in Fig. 1(a) and 1(b). The plots show that the user check-in distribution in both datasets fits a truncated power law distribution, with exponent  $e = 2.11$  and  $1.36$ , and  $\lambda = 6.6e - 03$  and  $6.2e - 03$ , in GW and BK, respectively. Both distributions show that most users have a low number of check-ins, and few have a large number of check-ins. More specifically, 16% and 30% (22% and 35%) of users in respectively Gowalla and Brightkite data set have trajectories shorter than or equal to 3 (5). This observation highlights the need for learning from more than the single individual's trajectory.

#### C. Analysis of Links and Mobility

We investigate the correlation between *human mobility* and *online social links*, to motivate amplifying individuals' trajectories by those of their social ties. Precisely, we aim to answer the question: Is individuals' mobility correlated with their 1, 2, and 3-degree links?

To answer the question, we first propose an approach to quantify the impact of  $\mathbf{a}_v^\kappa$ 's mobility on  $v$ . To do so, we reconstruct  $v$ 's mobility vector given only the mobility of users in  $\mathbf{a}_v^\kappa$ . We then compute the *reconstruction sampling error* between a sample<sup>3</sup> from the original mobility vector of  $v$ ,  $\mathbf{x}_v$ , and the reconstructed trajectory sample,  $\hat{\mathbf{x}}_v$ . Then, we set  $\kappa = 1, 2, 3$  in  $\mathbf{a}_v^\kappa$  for including the 1-, 2-, and 3-hop connected neighbors.

<sup>3</sup>We sample to overcome the curse of dimensionality, where sampling is achieved by considering all locations visited by  $v$  in addition to a random subset of non-visited locations.

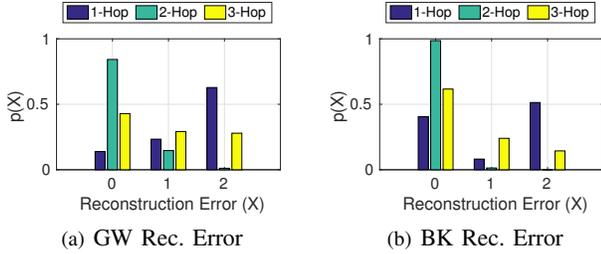


Figure 2. Reconstruction Error for [1, 2, 3]-Hops.

We report the distribution of reconstruction error for all settings in Fig. 2, where errors closer to zero indicate better reconstruction. Clearly, the results show that 1) individuals’ mobility is correlated with their social links and 2) higher correlation is found between mobility of individuals and those of their 2-hop social links. As a result, we propose a representation learning model that learns patterns from individuals along with their 2-hop social connections.

#### IV. MODELING HUMAN MOBILITY

##### A. STAR: The Proposed Model

We design an LDA-like [4] approach due to the observed similarities in distributions between human mobility data portrayed by LBSNs and text corpora, as well as to the success and popularity of topic models in text mining. Similar to the setting of topic models, we are given locations (documents), each of which has a set of users (words) who visited. Users visit the same location (words appearing in the same document) due to some *activities* (topics). *Activities* are the latent patterns to infer by the propose model. Each location is then represented as a distribution over activities,  $\theta_c \in \mathbb{R}_+^K$ ,  $c = 1 \dots |\mathcal{C}|$ , and each activity has an attractiveness distribution over users (participants),  $\phi_k \in \mathbb{R}_+^{|\mathcal{V}|}$ ,  $k = 1 \dots K$ . The contribution of social ties is added to the latent patterns using a generalized Pólya Urn sampling, with minimum cost (only a constant factor is added to the cost of regular LDA), and is scalable to  $|\mathcal{V}|$ , as the degree of user ties follows a heavy-tailed power law distribution (demonstration figure is omitted due to space limitation). Additionally, our model enforces an asymmetric prior over location distribution to handle the power law distribution of users check-ins.

To understand the representation learning process, we explain how a user  $v$  visits a location  $c$ . For one location  $c$ , it has a distribution over activities (e.g., a plaza can host a variety of events). Denoting this per-location distribution over activities  $\theta_c$ , it can be sampled from a Dirichlet distribution with an asymmetric prior,  $\alpha$ , which itself is drawn from a Gamma distribution with two parameters  $n, \mathbf{u}$ .

For one user  $v$ , whether or not he/she visits  $c$  depends on two factors:

**Factor 1:** which type of activity is occurring at  $c$ ; and

**Factor 2:** his/her preference of attending such an activity.

We see the Factor 1 firstly. As we know,  $\theta_c$  is the probability of different activities happening at  $c$ . We can select an activity  $z$  according to the categorical distribution parameterized by  $\theta_c$ . For example, let a sunset concert happen at the plaza. *Which user will come for the concert?* We see the Factor 2. Each activity has its attractiveness to different users, sampled from a Dirichlet distribution with a prior  $\beta$ . Activity  $z$  (e.g., a concert) has attractiveness given in  $\phi_z$  showing the probability of attracting each user. A user is more likely to be attracted if he/she is interested in it. Therefore, a user  $v$  is chosen to visit  $c$  according to the categorical distribution parameterized with  $\phi_z$ . Once a user  $v$  visits a location  $c$ , his/her probability of visiting the same location again increases, analogously to sampling from a Pólya Urn distribution.

In fact, another factor attracting a user to visit  $c$  is,

**Factor 3:** how many of his/her friends visit  $c$ .

For example, user  $v$  may join a concert if his/her friends are going for it. We re-consider the attractiveness distribution  $\phi_z$  of one activity over users. Suppose user  $v$  is socially connected with  $v'$  ( $a_{v,v'} = 1$ ). Due to the social influence we analyzed previously, the probability of user  $v$  going for  $z$  is increased if  $v'$  also goes for  $z$ . We thus should update  $\phi_z$  by including  $v'$  if  $a_{v,v'} = 1$ ,  $\forall v' \in \mathcal{V}$ , as given in Eq. (4), such that the element in  $\phi_z$  corresponding to  $v$  is amplified by that of socially connected  $v'$ . This sampling approach is known as generalized Pólya Urn, where the social ties,  $\mathbf{A}$ , are referred to as the addition matrix.

Comparing to basic LDA, our model has two advantages, 1) it utilizes a generalized Pólya Urn sampling approach for enhancing the activity attractiveness distributions over users based on their social ties; and 2) it adopts an asymmetric  $\alpha$  to handle the power law distribution of user check-ins.

The formal generative process goes as follows:

- 1) Draw  $\alpha \sim \text{Gamma}(n, \mathbf{u})$
- 2) For each *activity*,  $k = 1 : K$ 
  - a) Draw  $\phi_k \sim \text{Dirichlet}(\beta)$
- 3) For each *location*  $c \in \mathcal{C}$ 
  - a) Draw  $\theta_c \sim \text{Dirichlet}(\alpha)$
  - b) For each *user*  $v \in \mathcal{V}_c$ 
    - i) Choose  $z \sim \text{Categorical}(\theta_c)$
    - ii) Choose  $v \sim \text{Categorical}(\phi_z)$

where  $n, \mathbf{u}$  are the parameters of the Gamma distribution,  $\mathcal{V}_c$  is the set of users who have check-ins at  $c$  and  $z$  is an activity drawn from the location-activity distribution.

In STAR, only  $v$  is observed for forming  $\mathbf{X}$ , while all other variables are latent. We discuss the inference of these parameters in next section. Finally, the  $\theta_c, c = 1 \dots |\mathcal{C}|$  compose a  $|\mathcal{C}| \times K$  matrix of location-activity distributions,  $\Theta$ , which is the new representation of locations. The  $\phi_k, k = 1 \dots K$  can be packed row by row, resulting in a  $K \times |\mathcal{V}|$  matrix  $\Phi$ , which is an activity-user distribution matrix.  $\Phi^T$  will be the new representation of users.

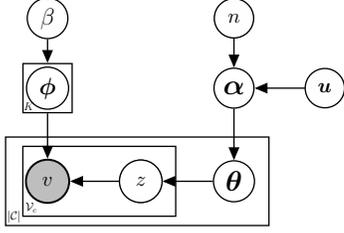


Figure 3. Graphical model of STAR. Nodes represent random variables, shaded nodes are observed, and plates indicate repetition.

## B. Model Parameter Estimation

Given the graphical model shown in Fig. 3, the posterior distribution of the latent variables  $(\mathbf{z}, \Phi, \Theta, \alpha)$  given the observed data is,

$$P(\alpha, \Phi, \Theta, \mathbf{z} | \mathbf{X}, n, \mathbf{u}, \beta) = \frac{P(\mathbf{X}, \alpha, \Phi, \Theta, \mathbf{z} | n, \mathbf{u}, \beta)}{P(\mathbf{X} | n, \mathbf{u}, \beta)}. \quad (1)$$

This distribution is intractable and cannot be computed exactly. Thus, we adopt an inference model by alternating between 1) estimating  $\mathbf{z}$  using Generalized Pólya Urn Gibbs sampling and 2) optimizing  $\alpha$  using fixed-point iteration.

1) *Gibbs Sampling*: Collapsed Gibbs sampling is an MCMC algorithm, which iteratively draws and updates one sample from the population given all other samples. In our model, this implies sampling from the conditional distribution,  $P(z_i | \mathbf{z}^{-i}, \mathbf{X})$ , since  $\mathbf{z}$  is a sufficient statistic for  $\Phi$  and  $\Theta$ . Sampling is given by,

$$P(z_i | \mathbf{z}^{-i}, \mathbf{X}) \propto (\eta_{z_i | \mathbf{x}}^{(-i)} + \alpha z) \cdot \frac{\sum_{v'} |\mathcal{V}| \eta_{v' | z}^{(-i)} a_{v, v'}^2 + \beta}{\eta_{\cdot | z}^{(-i)} + \beta}. \quad (2)$$

where  $\eta_{z_i | \mathbf{x}}$  is the number of times activity  $z$  was drawn when forming a check-in vector  $\mathbf{x}$  for one location, superscript  $(-i)$  excluded  $i$ -th element in  $\mathbf{z}$ ,  $\eta_{v | z}$  is the number of times user  $v$  attending activity  $z$ ,  $\eta_{\cdot | z} = \sum_{j=1}^{|\mathcal{V}|} \eta_{v_j | z}$  is the total number of users attending activity  $z$ , and  $\beta = \sum_{v=1}^{|\mathcal{V}|} \beta$ .

The activity-user distribution for each activity is  $P(\phi | \mathbf{X}, \beta) = \frac{1}{Z_\phi} \text{Dirichlet}(\phi | \eta_{\cdot | k} + \beta)$ , where  $Z_\phi$  is a normalization factor. The estimation of each element  $\phi_{v | z}$  in  $\phi$  should include the user adjacency matrix  $\mathbf{A}$ . We utilize the generalized Pólya Urn framework to the activity-user component of the model, in which we update each element in  $\phi_z$  by the sampled user along with his/her connections. This approach of boosting  $\phi$  by social ties, enforces coherence among users in the inferred activities, which serves the observation that mobility of connected users is correlated.

Since the expectation of  $\text{Dirichlet}(\mathbf{x} | \vartheta)$  is,

$$E(x_i) = \frac{\vartheta_i}{\sum_{k=1}^K \vartheta_k}, \quad (3)$$

one element of  $\phi$  (the attractiveness of activity  $z$  to user  $v$ )

is estimated using,

$$\phi_{v | z} = \frac{\sum_{v'} |\mathcal{V}| \eta_{v' | z} a_{v, v'}^2 + \beta}{\eta_{\cdot | z} + \beta}, \quad (4)$$

which considers the social ties of  $v$  (where  $a_{v, v'}^2 = 1$ ).

The per-location activity distribution is given by  $P(\theta | \mathbf{z}, \alpha) = \frac{1}{Z_\theta} \text{Dirichlet}(\theta | |\mathcal{V}_c| + \alpha)$ . Again, given Equation (3), the  $z$ -th element in the distribution vector  $\theta$  is,

$$\theta_{z | \mathbf{x}} = \frac{\eta_{z | \mathbf{x}} + \alpha z}{|\mathcal{V}_c| + \alpha}. \quad (5)$$

The model's complexity is equivalent to basic LDA, when  $\mathbf{A}$  is a sparse matrix, it is  $O(K \cdot |\mathcal{C}| \cdot |\mathcal{V}_c|)$ .

2) *Optimizing the Location-Activity Hyperprior*: Let the Dirichlet distribution parameter  $\alpha$  be decomposed into  $\alpha = pm$ , where  $\mathbf{m} = (m_1, m_2, \dots, m_K)$  is the base measure and  $p$  is the concentration parameter. Parameter  $pm_k$  is assumed to be independently drawn from the Gamma distribution,

$$P(pm_k | n, u) = \frac{1}{\Gamma(u)n} \left(\frac{pm_k}{n}\right)^{u-1} \exp\left(-\frac{pm_k}{n}\right). \quad (6)$$

Hyperprior  $\alpha$  can be estimated by maximizing the posterior probability.  $P(\alpha | \mathbf{X}) \propto P(\mathbf{X} | \alpha) P(\alpha)$ , where the likelihood  $P(\mathbf{X} | \alpha)$  and the hyperprior  $P(\alpha)$  can be transformed into log forms. The summation of log-likelihood and log-prior has a lower bound, which can be used to find  $\alpha^* = [pm_k]^*$ , thus maximizing  $P(\alpha | \mathbf{X})$ . The fixed-point iteration repeatedly runs Equation (7) and (8) until convergence is reached,

$$[pm_k]^{new} = [pm_k]^{old} \frac{\sum_{l=1}^L [\Psi(\eta_{k|l} + [pm_k]^{old}) - \Psi([pm_k]^{old})] + u}{\sum_{l=1}^L [\Psi(\eta_{\cdot | l} + [p]^{old}) - \Psi([p]^{old}) - \frac{1}{n}]}, \quad (7)$$

$$[p]^{new} = \sum_{k=1}^K [pm_k]^{old}, \quad (8)$$

where  $\Psi$  is the Digamma function (the first derivative of the log Gamma function).

## V. EVALUATION

**Experimental Design.** We evaluate the new user representation by measuring its performance on a link prediction task and compare the results to those of the representations learned using baseline models. We adopt two baseline models for comparisons: LDA [4], and LDA with asymmetric location-activity hyperprior (LDA-A) [8].

The evaluation framework has the following steps:

- 1) Given the *bag-of-users* location representation, we learn the new representations, location-by-activity and activity-by-user matrices, using either the proposed model or one of the baseline models stated above.
- 2) We renormalize the inferred activity-by-user matrix to get a user-by-activity representation, where each user is described as a distribution over inferred activities.
- 3) We apply unsupervised link prediction, detailed next, and compute the average AUC for measuring the quality

of link prediction. Higher AUC indicates more accurate link prediction and thus better representations<sup>4</sup>.

**Settings.** We apply all models to the same  $location \times user$  frequency matrix, set  $K$  to 5 for all graphical models and train them for a thousand iterations, and optimize parameters every 50 iterations, after an initial burn-in period of 200. The common parameter  $\beta$  for all graphical models is set to be  $\beta = 0.01$ . We study the model’s sensitivity to  $K$  after comparing the performance of different models.

### A. Representation Results

Fig. 4 shows heatmaps of the new representations learned by STAR. Fig. 4(a) and 4(b) show the per location activity representation,  $\theta_c$ , for a thousand randomly selected locations in GW and BK, respectively. Fig. 4(c) and 4(d) show the per user activity representation,  $\phi_v$ , for a thousand randomly selected users in GW and BK, respectively. We note the difference between  $\theta$  and  $\phi$ . The representations learned for users over activities are mostly sparse, indicating that users are mostly characterized by a single activity. However, there exist locations with uniform distribution over activities, e.g., the light lines from location 190 to location 200 in Fig. 4(a), though many locations are characterized by a single activity. Similar observations are obtained when varying  $K$ . Fig. 4(e) and 4(f) show the location and user representation of BK when  $K = 50$ . We see that a larger  $K$  reveals latent activities in a finer level. Both user and location representation heatmap show smaller and more clearly distinct block-diagonal pattern, without horizontal light lines. The distinct block-diagonal pattern of the heatmaps in Fig. 4 evidences the high quality of the obtained representation. It properly captures latent patterns of locations and users, and can instantly be used to group them into clusters.

We manually inspect 3 locations in GoogleMap and check their learned representations. Fig. 5(a) shows the actual locations in a map and the representations learned by STAR when  $K = 5$ . Location 85 and 22 have representations  $[0, 0, 0, 0, 1]$  and  $[1, 0, 0, 0, 0]$ , respectively. This indicates that location 85 is characterized by *Activity 5* only, and location 22 is characterized by *Activity 1* only. By zooming in the map, we see location 85 is a building of “The Plastics Industry Trade Association” and location 22 is Dunkin Donuts, in Fig. 5(b). It makes sense that these two locations are characterized by only one *Activity*, as users visit these places for simple and pure purpose, working or eating. Location 39 has a non-sparse representation  $[0.3, 0.3, 0.1, 0.1, 0]$ , which in other words is a mixture of different activities. By looking up, we see in Fig. 5(b) that it is a street place close to a bus station. It is checked-in maybe because people are waiting for bus there, or maybe because

<sup>4</sup>Note that we do not compare our link prediction results to those obtained by state of the art models, as our goal here is to evaluate the quality of the inferred patterns, which have proven to be very efficient when compared to unsupervised (topology) link prediction measures.

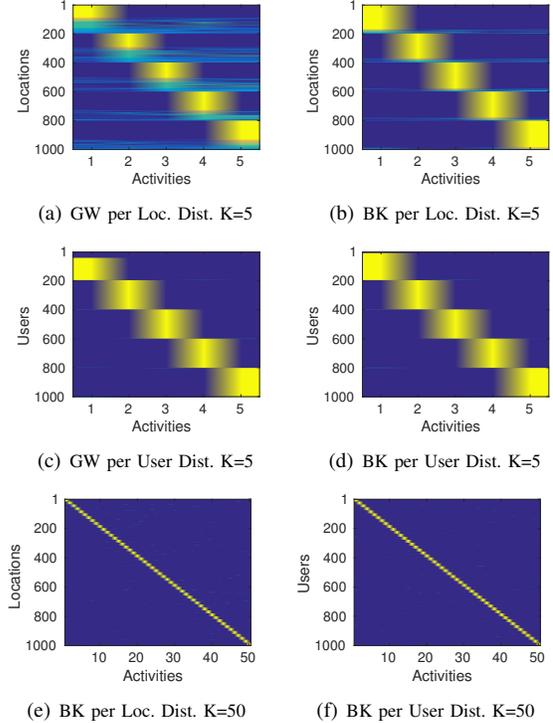


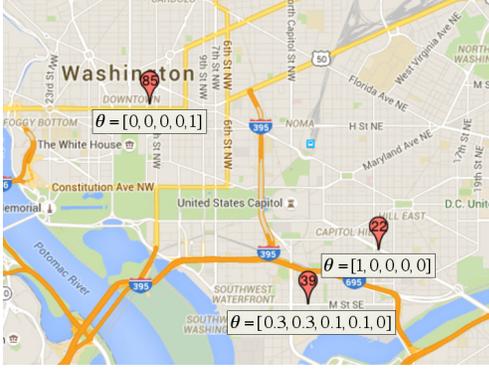
Figure 4. Heatmaps of per-location and per-user activity distributions achieved by STAR when  $K = 5$  and 50. The scale ranges from 0 (blue) to 1 (yellow).

people got stuck in their car due to traffic jams, or some other reasons. The learned representation characterizes well the possibility of different activities.

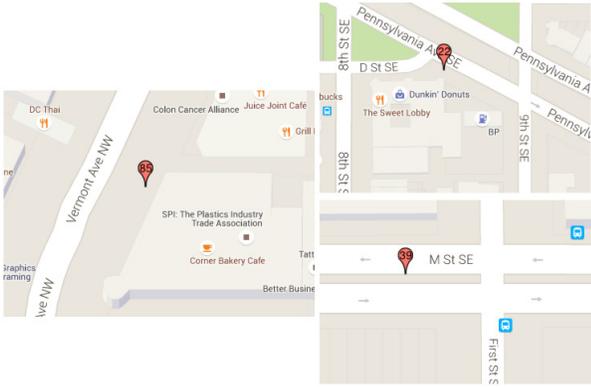
### B. Link Prediction

Given a subset of the network,  $A_{known}$ , which contains all nodes in  $A$  and only a subset of the edges, the task of link prediction is to infer the missing (new) links. Note that link prediction is a challenging problem due to the extremely unbalanced proportion of positive to negative links, where only 0.008% (0.015%) of elements in  $A$  are positive with value 1 in GW (BK). The study in [11] shows that topology measures computed from the user adjacency matrix  $A$  is able to predict links with good accuracy. Popularly used topology measures include Common Neighbors (CN), Jaccard Coefficient (Jacc), and Adamic-Adar (AA). A link is predicted between user  $v_i$  and  $v_j$  when their measure is above a threshold, varying which results in an ROC curve. The AUC value using each of these measures for both datasets is listed in Table II under the *Topology measure* column.

We employ the inferred user representations in unsupervised link prediction, to assess the quality of the representation. For one pair of users  $v_i$  and  $v_j$ , we calculate their proximity as  $-\|\phi_{v_i} - \phi_{v_j}\|_2^2$ , where  $\phi_v$  is user  $v$ ’s new representation. The existence of link between them is predicted when their proximity is above a threshold. Table II shows



(a) Three locations and their representations



(b) Zoom in the three locations

Figure 5. Case Study: Inspection of 3 locations

Table II AUC VALUES OF LINK PREDICTION

Data	Topology Measures			Learned Representations		
	CN	AA	Jacc	LDA	LDA-A	STAR
GW	0.592	0.557	0.526	0.669	0.596	<b>0.838</b>
BK	0.661	0.664	0.606	0.686	0.554	<b>0.840</b>

the AUC values when using new representations learned by LDA, LDA-A and STAR. In both datasets (GW and BK), the new representation learned by STAR outperforms all other representation learning methods and the topology measures.

Recall that both GW and BK have extremely sparse location-user frequency count matrix (about 0.003% non-zeros elements). STAR, unlike baselines, is shown to be able to capture patterns from the incomplete trajectories.

### C. Parameter Sensitivity

We study the sensitivity of STAR to the setting of the number of activities ( $K$ ). Fig. 6 shows the AUC results of link prediction for BK dataset when varying  $K$ . STAR significantly outperforms both LDA and LDA-A with much higher AUC values when varying  $K$ . Prediction results degrade when the number of activities increases over 50, in all models. As optimizing the location-activity hyper-prior minimizes the model’s sensitivity to  $K$ . This behavior is only reached when  $K$  is large enough, for small values of

$K$  such as in our experiments, a model with optimized  $\alpha$  is still sensitive to the user set parameter  $K$ .

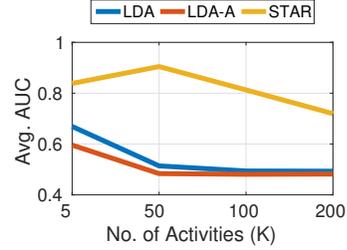


Figure 6. Sensitivity analysis of  $K$

## VI. CONCLUSIONS

We propose STAR, a PGM designed to infer latent patterns from anonymized and incomplete location-user traces, where user check-ins obey a power-law distribution. This is achieved by incorporating social links and asymmetric priors in the learning process, as a way of compensating for the limitations imposed by the incomplete data source.

## REFERENCES

- [1] N. Eagle and A. S. Pentland, “Eigenbehaviors: Identifying structure in routine,” *Behavioral Ecology and Sociobiology*, vol. 63, no. 7, 2009.
- [2] K. Farrahi and D. Gatica-Perez, “Discovering routines from large-scale human locations using probabilistic topic models,” *TIST*, vol. 2, no. 1, 2011.
- [3] B. Alharbi, A. Qahtan, and X. Zhang, “Minimizing user involvement for learning human mobility patterns from location traces,” in *AAAI*, 2016.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *JMLR*, vol. 3, 2003.
- [5] B. Kapicioglu, D. S. Rosenberg, R. E. Schapire, and T. Jebara, “Collaborative place models,” in *IJCAI*, 2015.
- [6] K. Joseph, C. H. Tan, and K. M. Carley, “Beyond local, categories and friends: clustering foursquare users with latent topics,” in *UbiComp*, 2012.
- [7] D. Phung, B. Adams, and S. Venkatesh, “Computable social patterns from sparse sensor data,” in *LocWeb*, 2008.
- [8] H. M. Wallach, D. M. Mimno, and A. McCallum, “Rethinking LDA: Why priors matter,” in *NIPS*, 2009.
- [9] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, “Optimizing semantic coherence in topic models,” in *EMNLP*, 2011.
- [10] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” <http://snap.stanford.edu/data>, Jun. 2014.
- [11] D. Liben-Nowell and J. Kleinberg, “The link prediction problem for social networks,” *Journal of the American society for information science and technology*, vol. 58, no. 7, pp. 1019–1031, 2007.