

Predicting neighbor label distributions in dynamic heterogeneous information networks

Yuchi Ma¹ · Ning Yang¹ · Lei Zhang¹ · Philip S. Yu^{2,3}

Received: 19 April 2016 / Revised: 13 September 2016 /
Accepted: 11 January 2017 / Published online: 19 January 2017
© Springer Science+Business Media New York 2017

Abstract In Dynamic Heterogeneous Information Networks (DHINs), predicting neighbor label distribution is important for a variety of applications. For example, when a user changes job, the composition of the user’s friends can change, hence the profession distribution of his/her social circle may change. If we can accurately predict the change of the distribution, we will be able to improve the quality of personal services for him/her. The challenges of predicting neighbor label distribution mainly come from four aspects: infinite state space of neighbor label distributions, link sparsity, the complexity of link formation preferences, and the stream of DHIN snapshots. To address these challenges, we propose a Latent Space Evolution Model (LSEM) for the prediction of neighbor label distribution, which builds a Neighbor Label Distribution Matrix (NLDM) for each type of labels of neighbors of given nodes. LSEM can predict the next NLDM by reconstructing it from two latent feature matrices estimated by their respective autoregressive models. The experiments conducted on real datasets verify the effectiveness of LSEM and the efficiency of the proposed algorithm.

Yuchi Ma and Ning Yang are equally contributed to this study and share the first authorship.

✉ Yuchi Ma
scu.Richard.Ma@gmail.com

✉ Ning Yang
yangning@scu.edu.cn

Lei Zhang
Leizhang@scu.edu.cn

Philip S. Yu
psyu@uic.edu

¹ College of Computer Science, Sichuan University, Chengdu, China

² Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

³ Institute for Data Science, Tsinghua University, Beijing, China

Keywords Dynamic heterogeneous information network · Neighbor label distribution prediction · Latent space evolution model

1 Introduction

In recent years, prediction in dynamic heterogeneous information networks (DHINs) has become a hot topic owing to the development of social networks (e.g., Facebook, Twitter and Foursquare). Existing works mainly focus on predicting the rating [7, 26, 29, 32] or the formation [9, 18, 19, 38, 41] of a single link. Some other works focus on predicting the labels of a single node [1, 3, 8, 15, 48]. In this paper, we aim to address the problem of the prediction of the neighbor label distributions of given nodes in a DHIN.

In a DHIN, a node might have neighbors with different types in different periods. For example, Figure 1 shows a stream of co-author network snapshots, where the nodes and labels are represented by circles and rectangles, respectively. As we can see from Figure 1, there are two types of nodes, Author nodes and Paper nodes, and each type of nodes may have multiple types of labels. For example, in Figure 1, an Author node has a Profession label (e.g., the label "Student" of Author 1), while a Paper node has a domain label (e.g., the label "Database" of Paper 1) and a publishing category label (e.g., the label "Journal" of Paper 1). For a given node, we can build a distribution for each type of the labels of its neighbor nodes, which we call neighbor label distribution. For example, there are three types of neighbor label distributions of Author 1 in Figure 1, where each type of neighbor label distribution is represented by a 100 % stacked as shown in Figure 2. As we can see from Figure 2, the neighbor label distributions are time-evolving.

Predicting neighbor label distributions is valuable for a variety of applications. For example, it is important for a scholar to keep up with the academic frontier by predicting the distributions of the research interests of leading scholars and following them. For more examples, a location-based social network may want to predict regions a user will go to tomorrow so as to recommend venues for the user. A movie rating network may want to predict the spending power distribution of the potential consumers on a new TV play for making a proper price of the play. A social media network may want to predict the interested topic distribution of each user so as to design his/her personalized start page for each log-in.

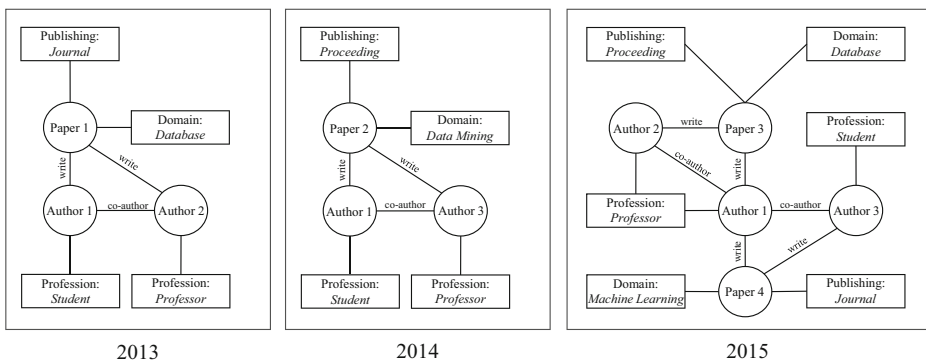


Figure 1 A stream of co-author network snapshots

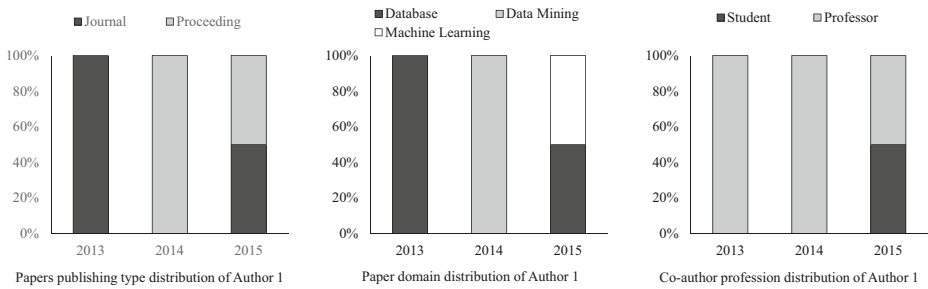


Figure 2 Neighbor label distributions

However, predicting neighbor label distributions is not easy due to the following challenges.

- Infinite state space of neighbor label distributions:** The number of possible states of a neighbor label distribution is infinite, because fractions in a neighbor label distribution are real numbers. Temporal models such as Markov chain-based models and time varying relational classifier [8, 33] cannot serve our goal because they often assume a finite state space.
- Sparsity of links:** In real-world networks, the number of links between a given node and its neighbors is relatively small compared with the huge volume of nodes, e.g., "publishing" in DBLP and "checking in" in Foursquare. The sparsity of links makes it difficult to mine sufficient meaningful patterns for individuals.
- The complexity of link formation preference:** A neighbor label distribution essentially reflects the link formation preference of a given node to the nodes with a specific type, and the evolution of the distribution results from the preference shifting. For example, in a co-author network, the paper domain distribution of an author changes with his/her research interests. Some previous researches [5, 17] indicate that the link formation preference might be very complicated as it can be affected by multiple factors, e.g., social circle, profession, age, and so on. Therefore, it is difficult to accurately model the link formation preference of a node by using specific features.
- Stream of DHIN snapshots:** As we can see from Figure 1, neighbor label distributions of a given node evolve over a stream of DHIN snapshots. Sometimes we may want to get real-time response to each snapshot. For example, the recommender of a video website might want to predict the genre distribution of users' playlists so as to arrange personalized start pages for them in real time. Many traditional methods are not applicable to make a prediction by only using the latest information.

To address these challenges, we propose a Latent Space Evolution Model (LSEM) for the prediction of neighbor label distribution, based on which the time cost of the predictions is linear with the number of nodes and independent of the number of DHIN snapshots. The main idea of LSEM is to model neighbor label distributions of given nodes as a Neighbor Label Distribution Matrix (NLDM) and make prediction of next NLDM by autoregressions of latent features. We call the given nodes *source nodes* and the labels *target labels*. Figure 3 gives an overview of LSEM. In the first part, to deal with the link sparsity and the complexity of link formation preference, LSEM factorizes each historical NLDM into a Source Node latent feature Matrix (SNM) and a Target Label latent feature Matrix (TLM). A row vector of SNM implicitly represents the link formation preference of a source node, while a row

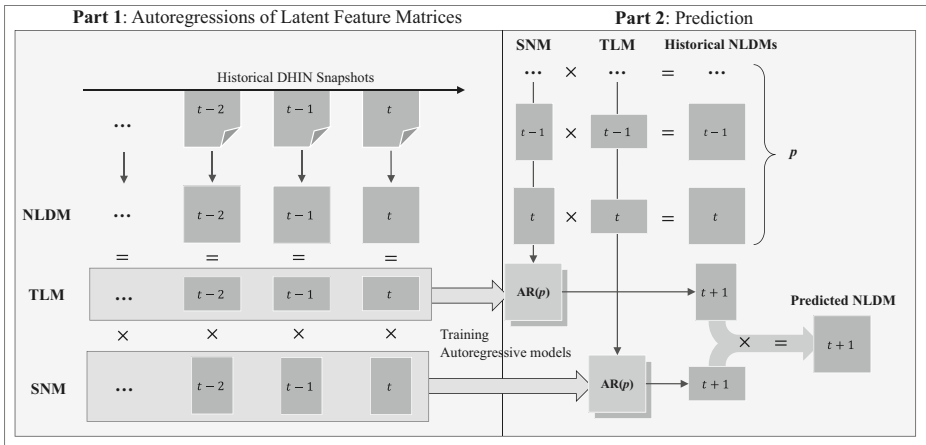


Figure 3 Overview of LSEM

vector of TLM represents the latent feature of a target label. By the factorization, we can model the link formation preference as a dense latent feature vector, even if the links around a node in a DHIN snapshot are sparse. To deal with the stream of DHIN snapshots, we fit the historical TLMs and SNMs by two p -order autoregressive models. As shown in the second part in Figure 3, LSEM makes a prediction of the next NLDM by reconstructing it from the next SNM and TLM which are estimated from the latest p historical SNMs and TLMs by the autoregressive models, respectively.

Our previous work [23] presents the problem of prediction of neighbor label distribution in an HIN, and proposes an Evolution Factor Model (EFM) for it. There are two major differences between LSEM and EFM. First, LSEM is an incremental model which makes a prediction based on the latest p snapshots, while EFM is a batch model which makes a prediction based on the whole history of an HIN. Second, LSEM takes into consideration the evolution of link formation preference of source nodes, as it is designed for a stream of DHIN snapshots, while EFM does not as it is for an HIN of which the history is treated holistically.

Our contributions can be summarized as follows:

1. We propose a Latent Space Evolution Model (LSEM) for the prediction of neighbor label distribution in DHINs. The time complexity of prediction algorithm based on LSEM is linear with the number of source nodes and independent of the number of DHIN snapshots.
2. We propose two p -order autoregressive models for the estimation of the next SNM and TLM, from which the NLDM can be reconstructed.
3. We compare LSEM with six baseline methods on four real datasets. The results verify the effectiveness and efficiency of the proposed model and algorithm.

The rest of this paper is organized as follows. We give the preliminaries in Section 2. We describe NLDM and formalize the problem of neighbor label distribution prediction in Section 3. We present the details of LSEM in Section 4. We present the experimental results and analysis in Section 5. Finally, we discuss related works in Section 6, and conclude in Section 7.

2 Preliminaries

In this paper, a vector is denoted by a boldface lowercase letter (e.g., \mathbf{v}), where the i th cell is denoted by $x^{(i)}$. A matrix is denoted by a boldface capital letter (e.g., \mathbf{X}), where the cell at i th row and j th column of \mathbf{X} is denoted by $x^{(i,j)}$.

We denote a heterogeneous information network by $G = \langle V, E \rangle$, where V is the node set and E is the link set. A Dynamic Heterogeneous Information Network (DHIN) can be segmented into a stream of DHIN snapshots denoted by $\mathcal{G} = \langle G_1, G_2, \dots, G_t, \dots \rangle$, where G_t represents the snapshot of G at time t [37]. For example, in the co-author network case, each G_t is a network comprised of all papers published in year t , as well as the authors linking to them. The neighbor label distribution is formally defined as follow:

Definition 1 (Neighbor Label Distribution): For a given node v of type V_1 , its neighbor label distribution over labels of type L of neighbor nodes of type V_2 is a vector

$$(x^{(1)}, \dots, x^{(j)}, \dots, x^{(M)}),$$

where M is the number of difference labels of label type L , and an entry $x^{(j)}$, $1 \leq j \leq M$, is the occurrence fraction of label j of type L , and $\sum_{j=1}^M x^{(j)} = 1$. The given node v is called **source node**, and the labels of type L are called **target labels**.

3 Neighbor label distribution matrix

For a given node, it has a neighbor label distribution uniquely corresponding to each possible combination of a neighbor node type and a neighbor label type. This corresponding relation can be modeled by a 2-hop meta-path [36] from a node type to a label type, which we call preference meta-path. The definition of the preference meta-path is given in Definition 2.

Definition 2 (Preference Meta-Path): A preference meta-path, denoted by \mathcal{R} , is a 2-hop relation from a node type to a label type,

$$\mathcal{R} = V_1 \rightarrow V_2 \rightarrow L, \quad (1)$$

where V_1 is a source node type, V_2 is a neighbor node type, and L is a label type.

A preference meta-path uniquely defines a type of neighbor label distributions. For example, in Figure 4, the paper domain distribution of authors can be defined by the

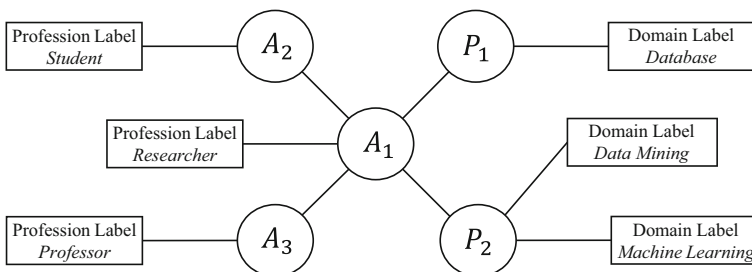


Figure 4 A snapshot of a co-author network

preference meta-path, $\mathcal{R} = Author \rightarrow Paper \rightarrow Domain$, and an instance of it is $A_1 \rightarrow P_2 \rightarrow Machine Learning$.

Based on the concept of preference meta-path, we propose Neighbor Label Distribution Matrix (NLDM) to model the neighbor label distributions of source nodes. The formal definition of NLDM is given as follow:

Definition 3 (Neighbor Label Distribution Matrix (NLDM)): The NLDM of a preference meta-path $\mathcal{R} = V_1 \rightarrow V_2 \rightarrow L$ at time t , denoted by $X_t(\mathcal{R})$, is a matrix of $\mathbb{R}^{N \times M}$, where N is the number of source nodes of type V_1 and M is the number of different labels of type L . The i th row vector in $X_t(\mathcal{R})$, denoted by $X_t^{(i)}(\mathcal{R})$, represents a neighbor label distribution, i.e.,

$$X_t^{(i)}(\mathcal{R}) = (x^{(i,1)}, \dots, x^{(i,j)}, \dots, x^{(i,M)}),$$

where $x^{(i,j)}$, $1 \leq j \leq M$, is the j th entry in $X_t^{(i)}(\mathcal{R})$. $x^{(j)}$ can be calculated as $x^{(j)} = \frac{PC^{i,j}(\mathcal{R})}{\sum_{l=1}^M PC^{i,l}(\mathcal{R})}$, where $PC^{i,j}(\mathcal{R})$ is the number of instances defined by the preference meta-path \mathcal{R} from the node i of type V_1 to the label j of label type L .

Hereinafter, for simplicity, we denote an NLDM by X_t and the i th row vector in X_t by $X_t^{(i)}$ in an unambiguous context.

As an illustration, Figure 5 shows two neighbor label distributions of node A_1 in the co-author network shown in Figure 4. Figure 5a shows the neighbor label distribution of A_1 defined by the preference meta-path $\mathcal{R}_1 = Author \rightarrow Author \rightarrow Profession$. Figure 5b shows the other neighbor label distribution of A_1 defined by the preference meta-path $\mathcal{R}_2 = Author \rightarrow Paper \rightarrow Domain$.

Now the problem of the prediction of neighbor label distributions can be reduced to the following prediction problem.

Problem statement Given a preference meta-path \mathcal{R} and the latest p NLDMs $\{X_{t-p+1}(\mathcal{R}), \dots, X_t(\mathcal{R})\}$ until current time t , we want to predict the NLDM at time $t + 1$, $X_{t+1}(\mathcal{R})$.

$\mathcal{R}_1 = Author \rightarrow Author \rightarrow Profession$

Profession	$PC^{A_1,j}(\mathcal{R}_1)$	$x^{(j)}$
Student ($j = 0$)	1	1/2
Professor ($j = 1$)	1	1/2
Researcher ($j = 2$)	0	0



$$(a) X_t^{(A_1)}(\mathcal{R}_1) = (\frac{1}{2}, \frac{1}{2}, 0)$$

$\mathcal{R}_2 = Author \rightarrow Paper \rightarrow Domain$

Domain	$PC^{A_1,j}(\mathcal{R}_2)$	$x^{(j)}$
Database ($j = 0$)	1	1/3
Data Mining ($j = 1$)	1	1/3
Machine Learning ($j = 2$)	1	1/3



$$(b) X_t^{(A_1)}(\mathcal{R}_2) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$$

Figure 5 Examples of neighbor label distribution vectors

4 Latent space evolution model

In this section, we present the details of Latent Space Evolution Model (LSEM), which consists of two parts, autoregressions of latent feature matrices and reconstruction of NLDM for prediction.

4.1 Autoregressions of latent feature matrices

For the prediction of Neighbor Label Distribution Matrix (NLDM), a straight forward idea is regression, however, which can not be directly applied due to the two challenges mentioned above, sparsity of links and the complexity of link formation preference. To overcome these issues, we first factorize each historical NLDM to latent feature matrices, and then utilizes two p -order autoregressive models to estimate the future latent matrices from which the future NLDM can be reconstructed.

We can factorize an NLDM X to a Source Node latent feature Matrix (SNM) and a Target Label latent feature Matrix (TLM) by solving the following optimization problem:

$$\operatorname{argmin}_{S, T} \frac{1}{2} \|X - S \times T^T\|_F^2 + \frac{\lambda}{2} (\|S\|_F^2 + \|T\|_F^2), \tag{2}$$

where $\|*\|_F$ denotes the Frobenius norm, $S \in \mathbb{R}^{N \times D}$ is the SNM, $T \in \mathbb{R}^{M \times D}$ is the TLM, N is the number of source nodes, M is the number of difference target labels, D is the dimensionality of latent features, λ is the balance parameter, and $\|S\|_F^2$ and $\|T\|_F^2$ are the regularization terms which avoid overfitting and guarantee the row vectors in S and T are dense.

A row vector of SNM implicitly represents the link formation preference of a source node, and a row vector of TLM represents the latent feature of a target label. Note that even if the links of a node in a DHIN snapshot are sparse, the latent features of the source node and the target label can still make sense because the SNM and TLM would be dense due to the regularization terms. Similarly, we can generate the historical SNMs and TLMs as follows:

$$\begin{aligned} X_1 &= S_1 \times T_1^T, \\ &\dots \\ X_{t-1} &= S_{t-1} \times T_{t-1}^T, \\ X_t &= S_t \times T_t^T. \end{aligned}$$

We estimate the SNM \widehat{S}_{t+1} and the TLM \widehat{T}_{t+1} from the latest p historical SNMs and TLMs respectively by using the following p -order autoregressive models:

$$\widehat{S}_{t+1} = \sum_{j=1}^p \phi_S^{(j)} S_{t-j+1} + C_S, \tag{3}$$

$$\widehat{T}_{t+1} = \sum_{j=1}^p \phi_T^{(j)} T_{t-j+1} + C_T, \tag{4}$$

where p is the order of the two autoregressive models, C_S and C_T are constant matrices in which all the entries are filled with $\phi_S^{(0)}$ and $\phi_T^{(0)}$, respectively, and $\{\phi_S^{(0)}, \dots, \phi_S^{(p)}\}$

and $\{\phi_{\text{T}}^{(0)}, \dots, \phi_{\text{T}}^{(p)}\}$ are parameters of the two autoregressive models. We learn the two autoregressive models by solving the following optimization problems:

$$\operatorname{argmin}_{\phi_{\text{S}}^{(0)}, \dots, \phi_{\text{S}}^{(p)}} \frac{1}{2} \sum_{t'=p+1}^t \|\mathbf{S}_{t'} - \widehat{\mathbf{S}}_{t'}\|_{\text{F}}^2, \quad (5)$$

$$\operatorname{argmin}_{\phi_{\text{T}}^{(0)}, \dots, \phi_{\text{T}}^{(p)}} \frac{1}{2} \sum_{t'=p+1}^t \|\mathbf{T}_{t'} - \widehat{\mathbf{T}}_{t'}\|_{\text{F}}^2, \quad (6)$$

where $\widehat{\mathbf{S}}_{t'} = \sum_{j=1}^p \phi_j \mathbf{S}_{t'-j} + \mathbf{C}_{\text{S}}$ and $\widehat{\mathbf{T}}_{t'} = \sum_{j=1}^p \phi_j \mathbf{T}_{t'-j} + \mathbf{C}_{\text{T}}$ are the estimated SNM and TLM at time t' , respectively. The order of the autoregressive models can be determined by minimizing the following residual [4]:

$$e = \sum_{t'=p+1}^t (\|\mathbf{S}_{t'} - \widehat{\mathbf{S}}_{t'}\|_{\text{F}} + \|\mathbf{T}_{t'} - \widehat{\mathbf{T}}_{t'}\|_{\text{F}}), \quad (7)$$

where $\widehat{\mathbf{S}}_{t'} = \sum_{j=1}^p \phi_j \mathbf{S}_{t'-j} + \mathbf{C}_{\text{S}}$ and $\widehat{\mathbf{T}}_{t'} = \sum_{j=1}^p \phi_j \mathbf{T}_{t'-j} + \mathbf{C}_{\text{T}}$ are the estimated SNM and TLM at time t' , respectively.

Note that in (5) – (7), p decides how many historical snapshots will impact the current SNM and TLM. When the setting value of p is too large, the earlier historical snapshots become disturbance items, which will result in the increasing of the residual e . On the contrary, when the setting value of p is too small, the model will be underfitting, which will also increase the residual e . Therefore, the optimal p depends on datasets and is experimentally determined by minimizing the residual e defined by (7). In our paper, we determine p based on experimental result as shown in Figure 11. For the datasets used for our experiments, p is chosen as 2 for DBLP, Foursquare and Yelp, and 3 for Netflix.

Algorithm 1 gives the learning algorithm of autoregressive models.

Algorithm 1 Learning Algorithm of Autoregressive Models ($\{\mathbf{X}_1, \dots, \mathbf{X}_t\}, D$)

INPUT:

$\{\mathbf{X}_1, \dots, \mathbf{X}_t\}$: The set of NLDMs at $\{1, \dots, t\}$;

D : The dimensionality of latent features;

OUTPUT:

$\Phi_{\text{S}} = \{\phi_{\text{S}}^{(0)}, \phi_{\text{S}}^{(1)}, \dots, \phi_{\text{S}}^{(p)}\}$: The set of parameters in the autoregressive model of SNM;

$\Phi_{\text{T}} = \{\phi_{\text{T}}^{(0)}, \phi_{\text{T}}^{(1)}, \dots, \phi_{\text{T}}^{(p)}\}$: The set of parameters in the autoregressive model of TLM;

p : the order of the autoregressive models of both SNM and TLM;

- 1: **for** $i = 1 \rightarrow t$ **do**
 - 2: Factorize \mathbf{X}_i into \mathbf{S}_i and \mathbf{T}_{ui} ;
 - 3: **end for**
 - 4: Determine p by minimizing the residual defined by (7);
 - 5: Apply gradient descent method to learn Φ_{S} of the autoregressive model defined by (3) by minimizing the objective function defined by (5);
 - 6: Apply gradient descent method to learn Φ_{T} of the autoregressive model defined by (4) by minimizing the objective function defined by (6);
-

4.2 Prediction

After the parameters, p , Φ_S and Φ_T , are learnt, we can predict X_{t+1} from the given p latest NLDMs, $\{X_{t-p+1}, \dots, X_{t-1}, X_t\}$. We first generate the SNMs $\{S_{t-p+1}, \dots, S_{t-1}, S_t\}$ and TLMs $\{T_{t-p+1}, \dots, T_{t-1}, T_t\}$ from the latest p historical NLDMs. Then, we estimate \widehat{S}_{t+1} and \widehat{T}_{t+1} by using the two learnt autoregressive models defined by (3) and (4). Finally, we can make a prediction of NLDM \widehat{X}_{t+1} as

$$\widehat{X}_{t+1} = \widehat{S}_{t+1} \times \widehat{T}_{t+1}^T, \quad (8)$$

where \widehat{S}_{t+1} and \widehat{T}_{t+1} are the estimated SNM and TLM.

Algorithm 2 gives the neighbor label distribution prediction algorithm. The time complexity of matrix factorization (Line (2)) is $O(DET)$, where E is the number of entries in X_t , D is the dimensionality of latent features and T is the fixed maximum iteration times. Because D , T and p are constants, the time complexity of Line (1) to Line (3) is $O(E)$. The time complexity of autoregressive model (Line (4) and Line (5)) is $O(E)$. Thus, the overall time complexity of Algorithm 2 is $O(E)$, which indicates the time complexity of Algorithm 2 is linear with the number of source nodes and independent of the number of DHIN snapshots.

Algorithm 2 Neighbor Label Distribution Prediction Algorithm ($\{X_{t-p+1}, \dots, X_t\}, D$)

INPUT:

$\{X_{t-p+1}, \dots, X_t\}$: The latest p NLDMs;

OUTPUT:

\widehat{X}_{t+1} : The predicted NLDM at time $t + 1$;

- 1: **for** $i = k - p + 1 \rightarrow k$ **do**
 - 2: Factorize X_t into S_t and T_t ;
 - 3: **end for**
 - 4: Estimate \widehat{S}_{t+1} by the autoregressive model defined by (3);
 - 5: Estimate \widehat{T}_{t+1} by the autoregressive model defined by (4);
 - 6: Predict \widehat{X}_{t+1} according to (8);
-

5 Experiments

In this section, we present the details of the experiments conducted on real datasets. We first determine the dimensionality D of latent feature and the order p of the autoregressive models, then verify the effectiveness and efficiency of LSEM. The experiments are executed on a Windows 7 PC with an Intel Xeon CPU of 3.4GHz and 16 GB RAM, and all programs are implemented in C#. The source code is available on [21].

5.1 Datasets

We conduct our experiments on four real datasets, two location-based social networks (Foursquare and Yelp), a co-author network (DBLP), and a movie rating network (Netflix), which are summarized in Table 1.

Table 1 Summary of datasets

Dataset	Preference meta-path			#Snapshot	#Latent feature
	V_1	V_2	L		
DBLP	#Author 13,153	#Paper 1,202,530	#Domain 25	14	11
Foursquare	#User 5,317	#Venue 46,065	#Region 8	24	22
Netflix	#User 21,439	#Movie 17,770	#Genre 28	24	6
Yelp	#User 11,214	#Business 57,476	#Category 14	28	9

5.1.1 DBLP

DBLP dataset is collected and maintained by the DBLP team [39]. The dataset indexes more than 2,800,000 articles.

On DBLP, given a preference meta-path $\mathcal{R} = Author \rightarrow Paper \rightarrow Domain$ and a group of authors, we want to predict the paper domain distributions of the authors. There are 25 domain labels on Computer Science. To obtain the domain labels of these papers, we first divide the titles of papers into keywords by using "NLTK" [27], which is a toolkit for building Python programs to work with human language data. Then we build a dictionary [22] which maps keywords to domain labels. At last, we extract the domain labels from keywords based on the dictionary. We select 13,153 authors who published more than 50 papers, and use their 1,202,530 papers which were published from 2000 to 2011. We use a year as the unit of time, thus the dataset is segmented into 12 snapshots each of which corresponds to one year of 2000 to 2011. Figure 6a shows that the number of authors follows a heavy-tailed distribution over the number of their published papers, which indicates the dataset is nature.

5.1.2 Foursquare

We joint two Foursquare datasets which are collected by Zhang et al. [14, 45, 46] and Bao et al. [2], respectively. One contains 5,392 users, 48,756 tips, 38,921 venues and 76,972 social links, and the other one contains 221,128 tips generated by 49,062 users at 46,065 venues in New York City (NYC).

On Foursquare, given a preference meta-path $\mathcal{R} = User \rightarrow Venue \rightarrow Region$ and a group of users, we want to predict the venue region distributions of the users' check-ins. We select 5,317 users who have more than 5 check-ins, and use their 59,498 check-ins in NYC from May. 1st, 2009 to May. 1st, 2011. We use a month as the unit of time, and segment the dataset into 24 snapshots. To obtain the region labels of venues, we divide NYC into 8 regions in terms of administrative divisions. Figure 6b shows that the number of users follows a heavy-tailed distribution over the number of check-ins, which again indicates the dataset is nature.

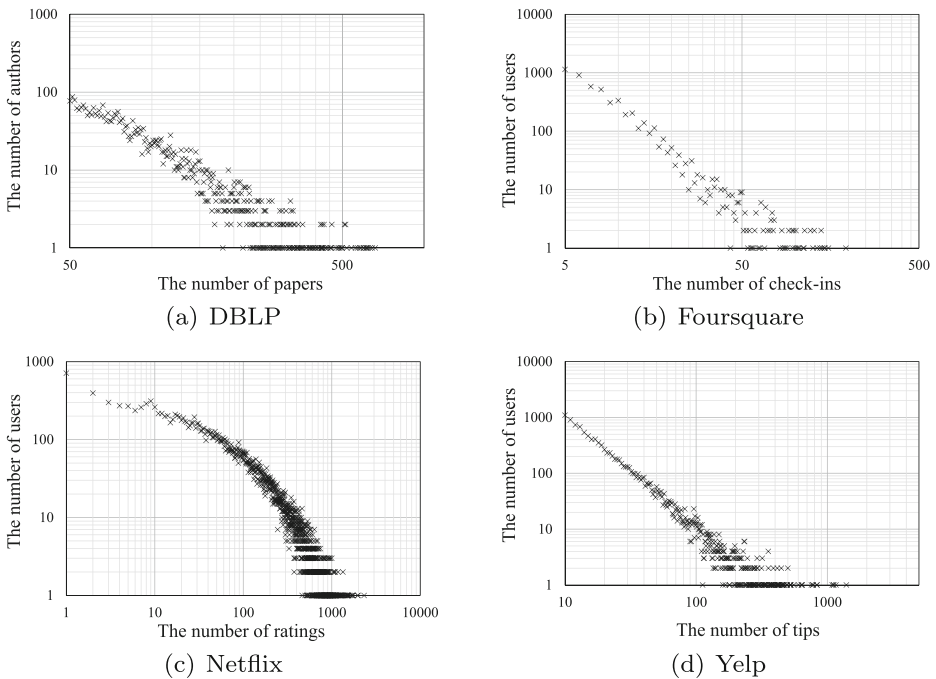


Figure 6 Overviews of datasets

5.1.3 Netflix

Netflix dataset [25] contains more than 100,000,000 rating records from about 480,000 customers over about 17,000 movie titles. The labels of "Movie" contain 28 genres crawled from the website IMDb [10].

On Netflix, given a preference meta-path $\mathcal{R} = User \rightarrow Movie \rightarrow Genre$ and a group of users, we want to predict the genre distributions of the users' rated movies. We select 21,439 users, and use their 2,704,432 ratings on 6,583 movies from Jan. 1st, 2000 to Jan. 1st, 2002. We use a month as the unit of time, and divide the dataset into 24 snapshots. Figure 6c shows that the number of users follows a heavy-tailed distribution over the number of movies, which also indicates the dataset is nature.

5.1.4 Yelp

Yelp dataset [43] contains more than 27,000,000 reviews and 649,000 tips by 687,000 users for 86,000 businesses. The labels of "Businesses" contain over 1000 categories, and we summarize these categories into 14 major categories.

On Yelp, given a preference meta-path $\mathcal{R} = User \rightarrow Business \rightarrow Category$ and a group of users, we want to predict the category distributions of the users' check-ins. We select 11,214 users who have more than 10 tips, and use their 648,303 tips from Oct. 1st, 2009 to Aug. 1st, 2016. We use a season as the unit of time, and segment the dataset into

28 snapshots. Figure 6d shows that the number of users follows a heavy-tailed distribution over the number of check-ins, which again indicates the dataset is nature.

5.2 Baseline methods

In order to demonstrate the effectiveness of our model, we compare LSEM with the following six baseline methods:

1. **Evolution Factor Model (EFM)**: An Evolution Factor Model (EFM) is proposed in [23] to infer neighbor label distribution in heterogeneous information networks. EFM considers a network itself as the historical network, and an update of the network as the current network. For a set of given nodes, EFM first clusters the nodes by using K -means, and then infers the neighbor label distribution for each cluster by directly applying a regression on the neighbor label distributions. To apply EFM for the prediction of NLDM, it takes the snapshot at time t as the current network, and the network comprised of all snapshots from time 1 to time $t - 1$ as the historical network. Then EFM is used to predict the NLDM at time $t + 1$.
2. **Matrix Factorization (MF)**: MF is proposed by B. Webb [42] to solve the movie recommender problem in Netflix Price. MF assumes the latent features of objects are modeled by vectors, and different types of objects have factors with the same size. When predicting the preferences of the objects of type A to the objects of type B , the preferences can be expressed as the product between the latent factor of the objects of type A and the latent factor of the objects of type B . To use MF for the prediction of neighbor label distribution, we build a historical NLDM $X = X_t + \dots + X_{t-p+1}$, and then factorize X by solving the following optimization problem:

$$\operatorname{argmin}_{S, T} \frac{1}{2} \|X - S \times T^T\|_F^2 + \frac{\lambda}{2} \left(\|S\|_F^2 + \|T\|_F^2 \right),$$

where λ is the balance parameter, S and T are the latent feature matrices of source nodes and target labels. We take the filled matrix $\hat{X} = S \times T^T$ as the predicted NLDM.

3. **Biased Matrix Factorization (Biased MF)**: Biased MF is proposed by Paterek [26], which is an extension of Matrix Factorization. Biased MF adds biased rates to the objects of either type. To apply Biased MF to the prediction of neighbor label distribution, we build a historical NLDM $X = X_t + \dots + X_{t-p+1}$, and then factorize X by solving the following optimization problem:

$$\operatorname{argmin}_{S, T, B_S, B_T} \frac{1}{2} \|X - S \times T^T - B_S - B_T\|_F^2 + \frac{\lambda}{2} \left(\|S\|_F^2 + \|T\|_F^2 \right),$$

where λ is the balance parameter, S and T are the latent feature matrices of source nodes and target labels, respectively, and B_S and B_T are the biased matrices of source nodes and target labels, respectively. We take the filled matrix $\hat{X} = S \times T^T$ as the predicted NLDM.

4. **Tucker Decomposition (TD)**: TD [40] decomposes a tensor into a product of a core tensor and matrices along all dimensions. To apply TD to prediction of neighbor label distribution, we construct the tensor and decompose it as shown in Fig. 7. The factor matrices can be computed by solving the following optimization problem:

$$\operatorname{argmin}_{C, S, T, K} \frac{1}{2} \|A - C \times_S S \times_T T \times_K K\|_F^2 + \frac{\lambda}{2} \left(\|S\|_F^2 + \|T\|_F^2 + \|K\|_F^2 \right),$$

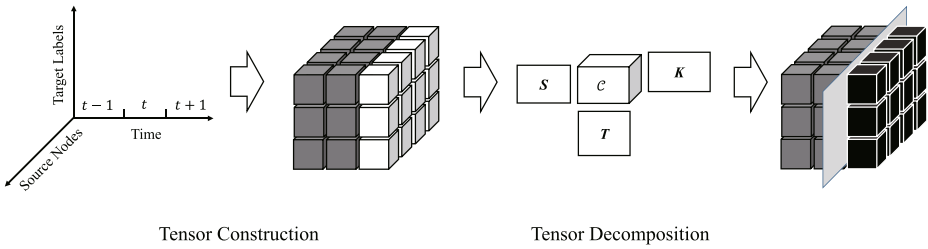


Figure 7 Tensor construction and decomposition for predicting neighbor label distribution

where \mathcal{A} is the constructed tensor shown in Figure 7, $\mathcal{C} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$ is the core tensor, $\mathbf{S} \in \mathbb{R}^{N \times D_1}$, $\mathbf{T} \in \mathbb{R}^{M \times D_2}$ and $\mathbf{K} \in \mathbb{R}^{3 \times D_3}$ are the latent feature matrices of source nodes, target labels and time, respectively, λ is the balance parameter, N is the number of source nodes, and M is the number of target labels. We take the sliced matrix at time $t + 1$ along the time dimension of the filled tensor $\hat{\mathcal{A}} = \mathcal{C} \times_S \mathbf{S} \times_T \mathbf{T} \times_K \mathbf{K}$ as the predicted NLDM.

- Context-Aware Tensor Decomposition (CATD):** CATD [47] decomposes a tensor into a product of a core tensor and factor matrices combining additional matrices built as context constraints. To apply CATD to the prediction of neighbor label distribution, we construct the same tensor as shown in the left part of Figure 7, and make a prediction of NLDM by solving the following optimization problem:

$$\begin{aligned} \operatorname{argmin}_{\mathcal{C}, \mathbf{S}, \mathbf{T}, \mathbf{K}} \quad & \frac{1}{2} \|\mathcal{A} - \mathcal{C} \times_S \mathbf{S} \times_T \mathbf{T} \times_K \mathbf{K}\|_F^2 + \frac{\lambda_1}{2} \sum_i \sum_j \left(\|S^{i*} - S^{j*}\|_F^2 S_s^{(ij)} \right) \\ & + \frac{\lambda_2}{2} \left(\|\mathbf{S}\|_F^2 + \|\mathbf{T}\|_F^2 + \|\mathbf{K}\|_F^2 \right), \end{aligned}$$

where S^{i*} is the latent feature vector of source node i , S_s is the relationship indicating matrix in which an entry $S_s^{(i,j)}$ is 1 if there is a link between i and j , and 0, otherwise. We take the sliced matrix at time $t + 1$ along the time dimension of the filled tensor $\hat{\mathcal{A}} = \mathcal{C} \times_S \mathbf{S} \times_T \mathbf{T} \times_K \mathbf{K}$ as the predicted NLDM.

- Meta Path-based Regression (MPR):** MPR [34, 35] uses a generalized linear model to build the connection between the observed building time of links and the meta path-based features. To apply MPR to the prediction of neighbor label distribution, we extract features based on two meta paths, "Author / User - Paper / Venue / Movie / Business - Domain / Region / Genre / Category" and "Author / User - Author / User - Paper / Venue / Movie / Business - Domain / Region / Genre / Category", and make a prediction of NLDM as follow:

$$\hat{x}_{t+1}^{(i,j)} = \exp\{-\mathbf{v}\boldsymbol{\beta}\},$$

where $\hat{x}_{t+1}^{(i,j)}$ is an entry of the predicted NLDM $\hat{\mathcal{X}}_{t+1}$ denoting author/user i and label j , \mathbf{v} is the meta path-based feature vector extracted from the historical snapshots, and $\boldsymbol{\beta}$ is the coefficient vector.

5.3 Metrics

The performance is measured by two metrics, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), defined as follows:

$$RMSE = \sqrt{\frac{\|\mathbf{X}_{t+1} - \widehat{\mathbf{X}}_{t+1}\|_F^2}{2N}},$$

$$MAE = \frac{\sum_{i=1}^N \|\mathbf{X}_{t+1}^{(i)} - \widehat{\mathbf{X}}_{t+1}^{(i)}\|_1}{2N},$$

where $\widehat{\mathbf{X}}_{t+1}$ is the predicted NLDM, \mathbf{X}_{t+1} is the ground truth, $\mathbf{X}_{t+1}^{(i)}$ and $\widehat{\mathbf{X}}_{t+1}^{(i)}$ are the i th row vectors in \mathbf{X}_{t+1} and $\widehat{\mathbf{X}}_{t+1}$, respectively, $\|\cdot\|_1$ denotes the L_1 -norm, and N is the number of source nodes.

5.4 Sensitivity of parameters

We now investigate the sensitivity of the parameters of LSEM, i.e., the dimensionality D of latent feature, the balance parameter λ , and the order p of the autoregressive models.

5.4.1 The dimensionality of latent features

For each of the four datasets, we tune the dimensionality of latent features, denoted by D , according to the residual of the matrix factorization on the first snapshot. The residual is calculated as

$$e = \|\mathbf{X}_1 - \mathbf{S}_1 \times \mathbf{T}_1^T\|_F,$$

where \mathbf{S}_1 and \mathbf{T}_1 are generated by solving the following optimization problem:

$$\operatorname{argmin}_{\mathbf{S}_1, \mathbf{T}_1} \frac{1}{2} \|\mathbf{X}_1 - \mathbf{S}_1 \times \mathbf{T}_1^T\|_F^2 + \frac{\lambda}{2} \left(\|\mathbf{S}_1\|_F^2 + \|\mathbf{T}_1\|_F^2 \right).$$

We tune D from 2 to 50, and as shown in Figure 8, the residuals on DBLP, Foursquare, Netflix and Yelp reach the minimum at 11, 22, 6 and 9, respectively. So we choose $D = 11$, $D = 22$, $D = 6$ and $D = 9$ as the dimensionality of latent features for the four datasets, respectively.

5.4.2 The balance parameter λ

For each of the four datasets, we determine the balance parameter λ , according to the performance of LSEM when setting different values of λ . We tune λ from 0.0 to 1.0 with step-size 0.05, and use the first 6 snapshots as the training set, and predict \mathbf{X}_{t+1} , $t = 6$ on DBLP. On Foursquare and Netflix, we use the first 18 snapshots as the training set, and predict \mathbf{X}_{t+1} , $t = 18$. On Yelp, we use the first 22 snapshots as the training set, and predict \mathbf{X}_{t+1} , $t = 22$. As shown in Figure 9, the RMSE and MAE of LSEM on DBLP, Foursquare, Netflix and Yelp reach the minimum at 0.15, 0.65, 0.2 and 0.6, respectively. So we choose $\lambda = 0.15$, $\lambda = 0.65$, $\lambda = 0.2$ and $\lambda = 0.6$ for the four datasets, respectively.

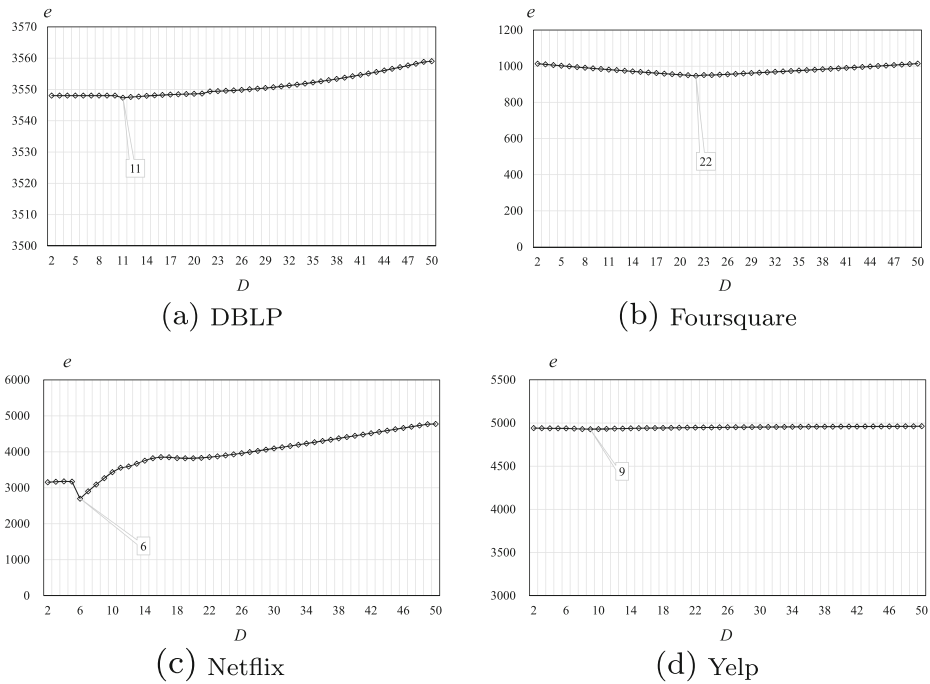


Figure 8 Tuning the dimensionality of latent features

5.4.3 The order p of autoregressive models

We first verify the stationarity of the sequences of latent features before applying autoregressive model for the estimation of the latent feature matrices, SNM and TLM. We randomly select 300 source nodes from each dataset and compute the autocorrelation of the sequence of each latent feature of each user in the first 12 snapshots of each dataset. For a sequence $\{s_1, s_2, \dots, s_t\}$, its autocorrelation r_q is calculated as:

$$r_q = \frac{\sum_{i=q+1}^t (s_i - \bar{s})(s_{i-q} - \bar{s})}{\sum_{i=1}^t (s_i - \bar{s})^2},$$

where q is the delay time, and \bar{s} is the average of the sequence. Then, for each of the latent feature matrices, SNM and TLM, we plot the average autocorrelation curves of each latent feature of each user in the same graph. As we can see from Figure 10, the autocorrelations all decline with q and drop to near zero, which shows that the sequences are stationary [4].

We then determine the order p according to residual analysis on snapshots of DBLP, Foursquare, Netflix and Yelp, respectively, where the residual is calculated as

$$e = \sum_{t'=p+1}^t (\|S_{t'} - \widehat{S}_{t'}\|_F + \|T_{t'} - \widehat{T}_{t'}\|_F).$$

We choose the order $p = 2$ for DBLP, Foursquare and Yelp, and $p = 3$ for Netflix, because the residuals on DBLP, Foursquare, Netflix and Yelp reach minimum at $p = 2$, $p = 2$, $p = 3$ and $p = 2$ respectively, as shown in Figure 11.

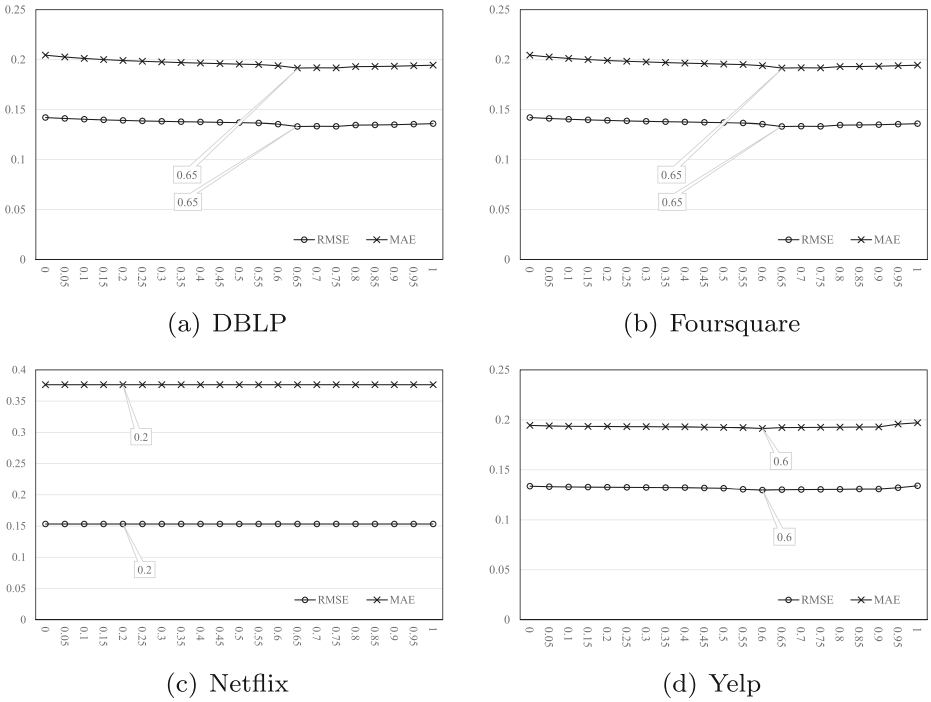


Figure 9 Tuning balance parameter

5.5 Effectiveness

On DBLP, we use the first 7 snapshots as the training set, and predict X_{t+1} , $7 \leq t \leq 11$. On Foursquare and Netflix, we use the first 19 snapshots as the training set, and predict X_{t+1} , $19 \leq t \leq 23$. On Yelp, we use the first 23 snapshots as the training set, and predict X_{t+1} , $23 \leq t \leq 27$. The results on the four datasets are shown in Figures 12, 13, 14 and 15 respectively.

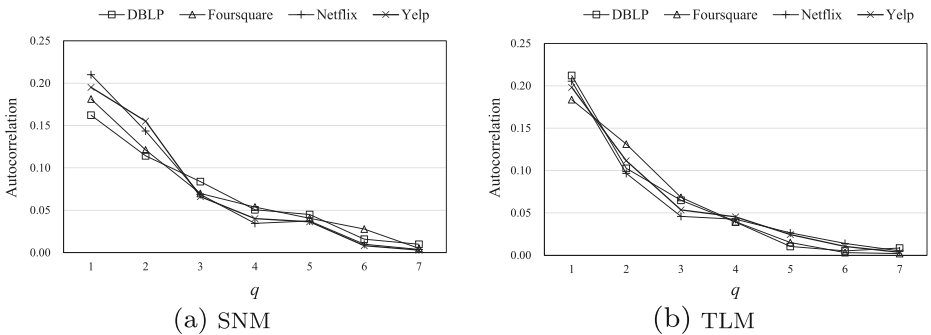


Figure 10 Average autocorrelation curves of DBLP, Foursquare, Netflix and Yelp

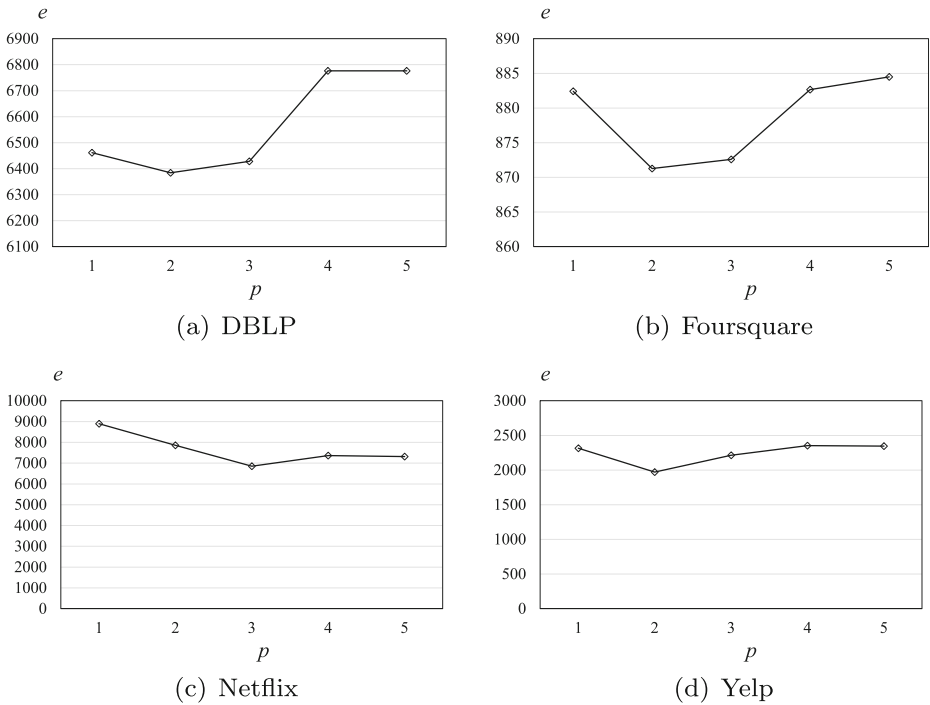


Figure 11 Residual of each p on DBLP, Foursquare, Netflix and Yelp

As we can see from the above figures, LSEM outperforms the baseline methods on all the datasets, of which the reasons can be analyzed as follows:

1. In contrast with LSEM, EFM makes predictions directly by a regression of NLDMs, which ignores link formation preference of source nodes.
2. MPR, MF, Biased MF, TD, and CATD can make predictions based on models incorporating the link formation preference, but they can not capture the evolution of link formation preference of source nodes.

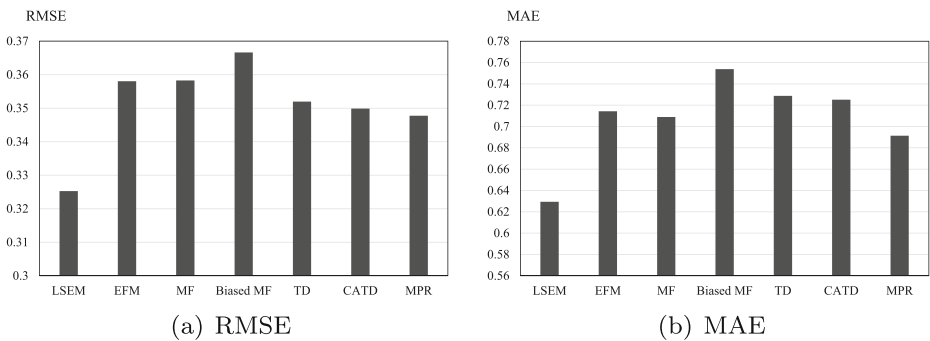


Figure 12 Performance comparison on DBLP

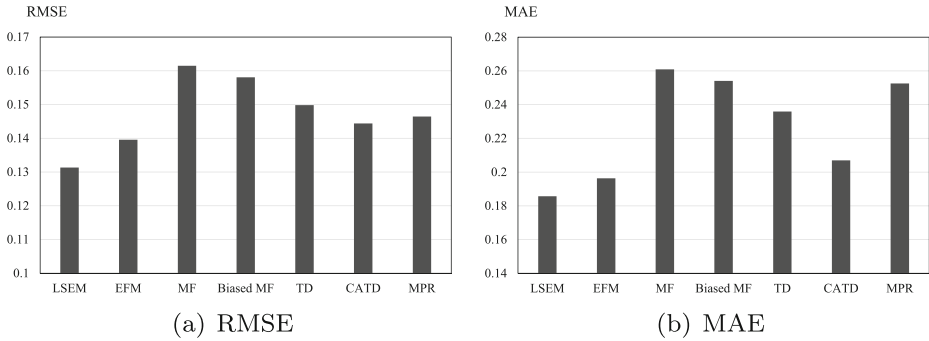


Figure 13 Performance comparison on Foursquare

5.6 Efficiency

As mentioned in Section 4.2, the time complexity of Neighbor Label Distribution Prediction Algorithm (NLDPA, shown in Algorithm 2) is linear with the number of source nodes and independent of the number of DHIN snapshots.

To verify the running time of NLDPA is linear with the number of source nodes, we randomly select 1000, 2000, 3000, 4000 and 5000 source nodes (Author nodes of DBLP, User nodes of Foursquare, Netflix, and Yelp) from each dataset, and run NLDPA for each group of source nodes. Figure 16 shows the running time of NLDPA is linear with the number of the source nodes on all the four datasets.

To verify the running time of NLDPA is independent of the number of DHIN snapshots, we use the first 7 snapshots as the training set for all the datasets. And we predict X_{t+1} , taking $7 \leq t \leq 11$ on DBLP, $7 \leq t \leq 23$ on Foursquare and Netflix, and $7 \leq t \leq 27$ on Yelp. As shown in Figure 17, the running time of prediction algorithm remains stable with the increasing number of DHIN snapshots on all the four datasets.

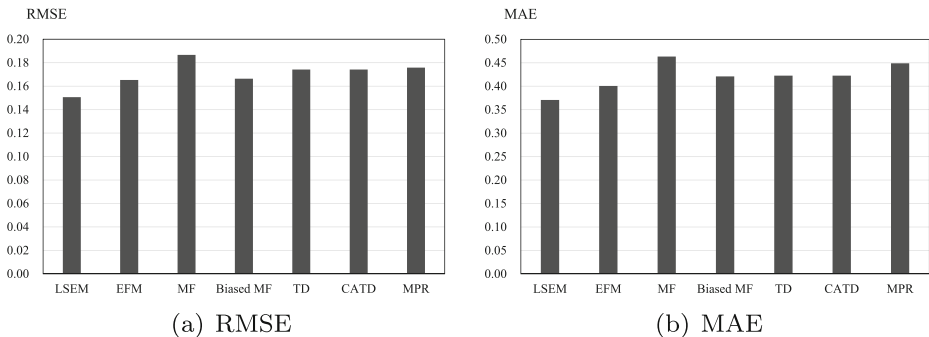


Figure 14 Performance comparison on Netflix

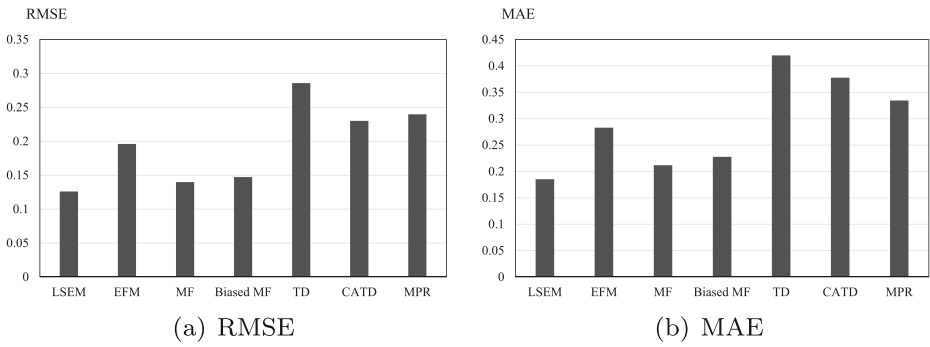


Figure 15 Performance comparison on Yelp

6 Related work

In this section, we first clarify the difference between this work and our previous work on the prediction of neighbor label distribution, then briefly introduce the related works from four domains, link prediction, recommending algorithms, user profiling, and node classification.

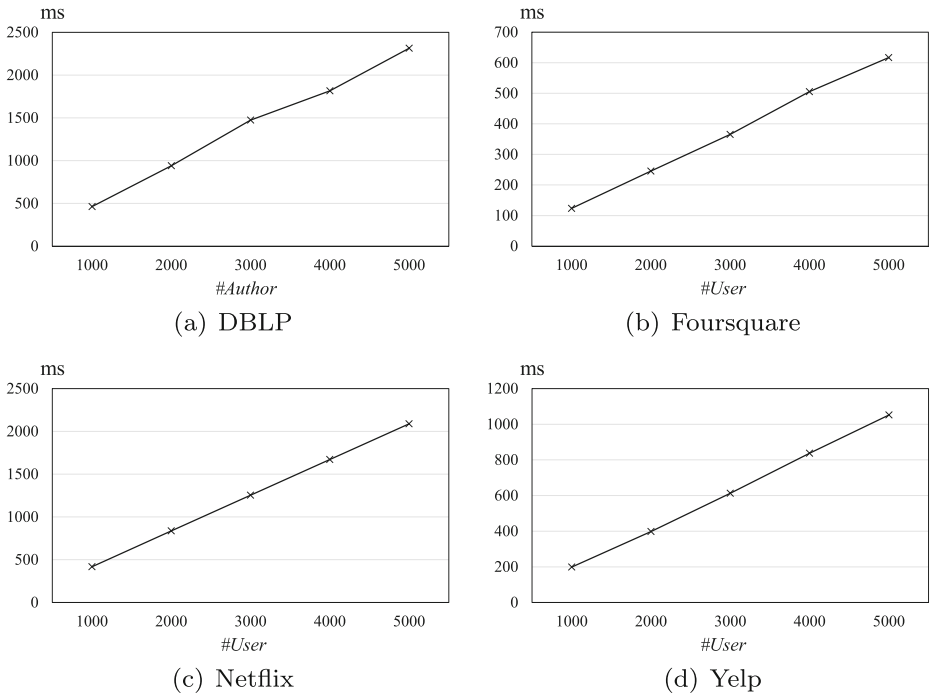


Figure 16 Running time of LSEM of different source nodes

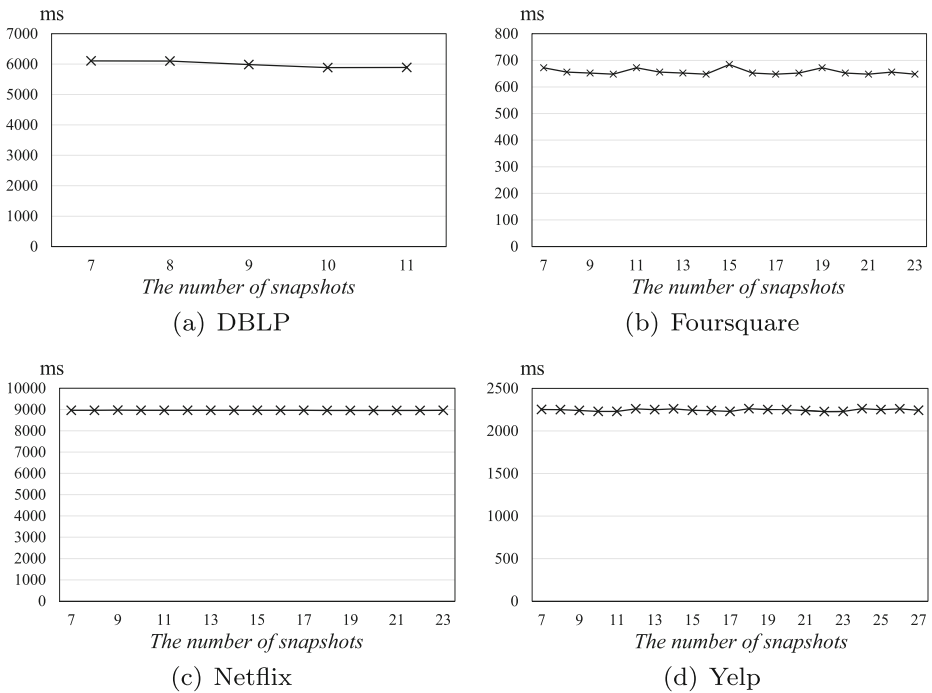


Figure 17 Running time of LSEM

Link prediction Hasan et al. [9] introduce a supervised learning method to predict whether a link will be built in the future. Wang et al. [41] propose a local probabilistic graphical model method that can scale to large graphs to estimate the joint co-occurrence probability of two nodes. Taskar et al. [38] propose a method to address the problem of link prediction in heterogeneous networks based on the observations of the attributes of the objects. There are a number of other studies which make use of the information about the change in the network over time for link prediction. For example, Zhu et al. [12] propose a Markov model-based method to predict links for helping user to surf the Internet, where the website page are clustered according to their hyperlinks. And the Markov model-based method is applied on the hierarchy constructed from the clustering results. Acar et al. [6] propose a time-aware link prediction method which constructs a three dimension tensor. The tensor structure includes the time information as the third dimension instead of collapsing the data, which overcomes the difficulty of representing temporal data. Sun et al. [35] extend the traditional link prediction to relationship prediction, which not only predicts whether it will happen, but also infers when it will happen. These methods mainly focus on the predictions around a single link, but pay little attention to the prediction of distributions.

Recommending algorithms Recommending algorithms can be classified as two categories: memory-based algorithms and model-based algorithms. The former type of methods make predictions based on homogeneous neighbors of given individuals directly [20, 28, 44], and the latter type of methods make recommendations based on a learned prediction model [29]. However, existing recommending algorithms mainly focus on predicting the rating of a single link between a user and an item, while in this paper, we aim at

the prediction of neighbor label distribution where the states of neighbors are considered holistically.

User profiling User profiling is an important task for many online services. Some early works focus on gender discrimination [16] and author identification [24]. Costa Jr et al. [13] propose an approach to predict demographic attributes and personalities based on the answers of users to some specific psychometric tests. Kosinski et al. [15] use dimensionality reduction for preprocessing the Likes data, which are then entered into logistic/linear regression to predict individual psychodemographic profiles. Brdar et al. [3] propose a multi-level classification model to predict users' genders, ages, marital statuses, job types, and so on. Lately, Zhong et al. [48] extract rich semantics of users' check-ins and propose the location to profile (L2P) framework to infer demographics of users. Our work is different from the user profile mining, because we focus on the dynamic characteristics of nodes which evolves over time.

Node classification Most node classification methods classify nodes in dynamic networks. For example, Sharan and Neville propose a relational Bayes classifier [30] and a probability tree [31] to classify nodes in an evolving heterogeneous network with different types of nodes, where time information is included to improve the classifier. Callut et al. [11] propose a novel technique, called *D*-walks, to tackle semi-supervised classification problems in dynamic network, which can deal with directed or undirected graphs with a linear time complexity with respect to the number of edges. Recently, a time varying relational classifier [8] is proposed to classify nodes in a heterogeneous network by effectively using nodes connected to the node and the connections from the past. However, these methods cannot serve our goal because they often assume a finite state space.

7 Conclusions

In this paper, we aim at addressing the problem of neighbor label distribution prediction in dynamic heterogeneous information networks. We propose a Latent Space Evolution Model (LSEM), which uses a Neighbor Label Distribution Matrix (NLDM) to represent the state of neighbors of given nodes. LSEM factorizes historical NLDM into Target Label latent feature Matrix (TLM) and Source Node latent feature Matrix (SNM) and predicts the next NLDM by reconstructing it from the next SNM and TLM which are estimated by the proposed autoregressive models. We conduct the experiments on four real datasets, two location-based social networks (Foursquare and Yelp), a co-author network (DBLP), and a movie rating network (Netflix). The results verify the effectiveness and efficiency of the proposed model and algorithm.

Acknowledgments This work is supported by National Science Foundation of China through grant 61173099, the Basic Research Program of Sichuan Province with Grant 2014JY0220, and US National Science Foundation through grants CNS-1115234, DBI-0960443, and OISE-1129076.

References

1. Back, M.D., Schmukle, S.C., Egloff, B.: How extraverted is honey.bunny77@hotmail.de? inferring personality from e-mail addresses. *J. Res. Pers.* **42**(4), 1116–1122 (2008)

2. Bao, J., Zheng, Y., Mokbel, M.F.: Location-based and preference-aware recommendation using sparse geo-social networking data. GIS '12 Proceedings of the 20th International Conference on Advances in Geographic Information Systems pp. 199–208 (2012)
3. Brdar, S., Culibrk, D., Crnojevic, V.: Demographic attributes prediction on the real-world mobile data. Proceedings of the Nokia mobile data challenge 2012 workshop (2012)
4. Cryer, J.D., Chan, K.S.: Time series analysis with applications in r, second edition (2008)
5. Diane, K., Teevan, J.: Implicit feedback for inferring user preference: a bibliography. ACM SIGIR Forum 37(2) (2003)
6. Acar, E., Dunlavy, D.M., Kolda, T.G.: Link prediction on evolving data using matrix and tensor factorizations. ICDM 2010 workshops. IEEE Computer Society pp. 262–269 (2010)
7. Gorrell, G., Webb, B.: Generalized hebbian algorithm for incremental latent semantic analysis. Proceedings of Interspeech (2006)
8. Gunes, I., Cataltepe, Z., Gunduz-Oguducu, S.: Ga-tvrc-het: genetic algorithm enhanced time varying relational classifier for evolving heterogeneous networks. Data Min. Knowl. Disc. **28**(2), 670–701 (2014)
9. Hasan, M.A., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning. SDM'06, Proceedings of SIAM International Conference on Data Mining (workshop on Link Analysis Counterterrorism and Security) (2006)
10. IMDb: <http://www.imdb.com/> (2015)
11. Callut, J., Francois, K., Saerens, M., Dupont, P.: Semi-supervised classification from discriminative random walks. ECML/PKDD **5211**, 162–177 (2008)
12. Zhu, J., Hong, J., Hughes, J.G.: Using markov models for web site link prediction. Hypertext. ACM pp. 169–170 (2002)
13. Costa, P.T. Jr., McCrae, R.R.: Reply to ben-porath and waller. Psychol. Assess. **4**(1), 20–22 (1992)
14. Kong, X., Zhang, J., Philip, Y.: Inferring anchor links across multiple heterogeneous social networks. CIKM '13 Proceedings of the 22nd ACM international conference on Conference on Information and Knowledge Management (2013)
15. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proc. Natl. Acad. Sci. **110**(15), 5802–5805 (2013)
16. Labov, W.: The Social Stratification of English in New York City. PhD thesis, Columbia University (1964)
17. Lamar, R.N., Beswick, R.W.: Voice mail versus conventional channels: A cost minimization analysis of individuals' preferences Academy of Management Journal 33(4) (1990)
18. Leroy, V., Cambazoglu, B.B., Bonchi, F.: Cold start link prediction. KDD'10, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (2010)
19. Lichtenwalter, R.N., Lussier, J.T., Chawla, N.V.: New perspectives and methods in link prediction. KDD'10. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining **26**(12), 2942–2955 (2010)
20. Linden, G., Smith, B., York, J.: Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing (2003)
21. Ma, Y.: Source code <https://pan.baidu.com/s/1gffUYmJ>
22. Ma, Y.: Keywords-direction dictionary. <http://pan.baidu.com/s/1kTmXd4N> (2015)
23. Ma, Y., Yang, N., Li, C., Zhang, L., Yu, P.: Predicting neighbor distribution in heterogeneous information networks SDM '14 (2014)
24. Mosteller, F., Wallace, D.L.: Inference in an authorship problem: a comparative study of discrimination methods applied to the authorship of the disputed federalist papers. J. Am. Stat. Assoc. **58**(302), 275–309 (1963)
25. Netflix prize. <http://www.netflixprize.com>
26. Paterek, A.: Improving Regularized Singular Value Decomposition for Collaborative Filtering. In: Proceedings of KDD Cup and Workshop 2007, vol. 2007, pp. 5–8 (2007)
27. Project, N.: Nltk 3.0 documentation. <http://www.nltk.org/> (2015)
28. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. WWW'10 Proceedings of the 19th international conference on World Wide Web (2001)
29. Savia, E., Puolamäki, K., Kaski, S.: Latent grouping models for user preference prediction. Mach. Learn. **74**(1), 75–109 (2009)
30. Sharan, U., Neville, J.: Exploiting time-varying relationships in statistical relational models. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis pp. 9–15 (2007)
31. Sharan, U., Neville, J.: Temporal relational classifiers for prediction in evolving domains. ICDM'08, Proceedings of the 8th IEEE International Conference on Data Mining pp. 540–549 (2008)

32. Shi, C., Li, Y., Yu, P.S., Wu, B.: Constrained-meta-path-based ranking in heterogeneous information network. *Knowledge Information System* pp. 1–29 (2016)
33. Sigal, S., Avi, R., Sarit, K., Navot, A.: A Hybrid Approach of Classifier and Clustering for Solving the Missing Node Problem. In: AAAI, pp. 282–289 (2015)
34. Sun, Y., Han, J.: *Mining heterogeneous information networks: Principles and methodologies* (2012)
35. Sun, Y., Han, J., Aggarwal, C.C., Chawla, N.V.: When will it happen? - relationship prediction in heterogeneous information networks. *WSDM'12 Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (2012)
36. Sun, Y., Han, J., Yan, X., Yu, P.S., Wu, T.: Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB'11* (2011)
37. Sun, Y., Tang, J., Han, J., Chen, C., Gupta, M.: Co-Evolution of Multi-Typed Objects in Dynamic Star Networks. *TKDE, IEEE Transaction on Knowledge and Data Engineering* (2013)
38. Taskar, B., fai Wong, M., Abbeel, P., Koller, D.: Link prediction in relational data. *Learning Statistical Patterns in Relational Data Using Probabilistic Relational Models 7* (2005)
39. Team, D.: Dblp. <http://dblp.uni-trier.de/> (2015)
40. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**, 279–311 (1966)
41. Wang, C., Satuluri, V., Parthasarathy, S.: Local probabilistic models for link prediction. *ICDM'07, Proceedings of IEEE International Conference on Data Mining 2007* (2007)
42. Webb, B.: Netflix update: Try this at home <http://sifter.org/~simon/JOURNAL/20061211.html> (2006)
43. Yelp dataset challenge. https://www.yelp.com/dataset_challenge
44. Yu, K., Schwaighofer, A., Tresp, V., Xu, X., Kriegel, H.P.: Probabilistic memorybased collaborative filtering. *TKDE Trans. Knowl. Data Eng.* **16**(1), 56–69 (2004)
45. Zhang, J., Kong, X., Philip, Y.S.: Predicting social links for new users across aligned heterogeneous social networks. *ICDM '13 Proceedings of the 13th International Conference on Data Mining* pp. 1289–1294 (2013)
46. Zhang, J., Kong, X., Philip, Y.: Transferring heterogeneous links across location-based social networks *WSDM '14 Proceedings of the 7th ACM international conference on Web search and Data Mining* (2014)
47. Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., Chang, E.: Diagnosing new york city's noises with ubiquitous data. *UbiComp '14, Proceedings of ACM International Conference on Ubiquitous Computing 2014* (2014)
48. Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F., Xie, X.: You are where you go: Inferring demographic attributes from location check-ins. *WSDM'15 Proceedings of the 8th ACM international conference on Web search and Data Mining* (2015)