This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

Social Recommendation with Cross-Domain Transferable Knowledge

Meng Jiang, Peng Cui, Xumin Chen, Fei Wang,
Wenwu Zhu *Fellow, IEEE*, and Shiqiang Yang *Senior Member, IEEE*

Abstract—Recommender systems can suffer from data sparsity and cold start issues. However, social networks, which enable users to build relationships and create different types of items, present an unprecedented opportunity to alleviate these issues. In this paper, we represent a social network as a star-structured hybrid graph centered on a social domain, which connects with other item domains. With this innovative representation, useful knowledge from an auxiliary domain can be transferred through the social domain to a target domain. Various factors of item transferability, including popularity and behavioral consistency, are determined. We propose a novel *Hybrid Random Walk* (HRW) method, which incorporates such factors, to select transferable items in auxiliary domains, bridge cross-domain knowledge with the social domain, and accurately predict user-item links in a target domain. Extensive experiments on a real social dataset demonstrate that HRW significantly outperforms existing approaches.

Index Terms—Social Recommendation, Transferability, Cross-Domain, Star-Structured Graph, Random Walk

1 INTRODUCTION

A social networking service is a platform on which users can create and adopt different types of items such as web posts (e.g., articles and tweets), user labels, images, and videos. The huge volume of items generates a problem of information overload. Traditional web post recommendation approaches suffer from data sparsity (i.e., limited interaction between users and web posts) and the issue of cold start (i.e., giving recommendations to new users who have not yet created any web posts). The social connections and multiple item domains found in social networks provide an unprecedented opportunity to alleviate these issues in real applications.

One common type of approach to recommendations, known as collaborative filtering techniques, characterizes users' latent features independently with user-item interactions in a single item domain [1]. Similarly, the type of approach provided in [2] does not consider the question of multiple domains. However, users' characteristics relate both to social connections and to different user-item interactions. For example, users read web posts created by their community and may adopt similar user labels to their friends. Therefore, an effective social recommendation approach should acknowledge (1) social tie strength (henceforth, tie strength) between users and (2) different user-item interactions. The problem of how to incorporate a social domain and auxiliary item domains (e.g., user labels and images) into a unified framework remains open.

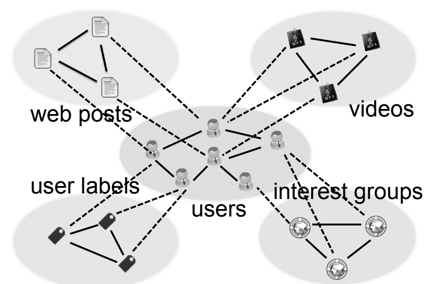


Fig. 1. Star-structured representation of a social network connecting multiple item domains and one social domain.

Frameworks exist that connect directly-related item domains, such as a music album and tags on that album [3], or web pages and queries to them [4]. However, these cannot be applied to indirectly-related item domains in social networks, such as tweets and user labels. The multiple item domains reflect users' intrinsic preferences and tend to be tightly connected among a massive number of users. In this paper, we reconsider the representation of social networks and propose a star-structured graph, where the social domain is at the center and is connected to the surrounding item domains, as shown in Figure 1.

The value of the *cross-domain link*[1] weight represents how often a given user adopts a given item, while the value of the *within-domain link*[2] weight in the social domain represents the tie strength between users. Tie strength can refer to homophily [5], circle-based influence [6][7][8], or social trust [9][10]. Users are more likely to have stronger ties if they share similar characteristics.

1. Cross-domain links are user-item links (item adoptions), i.e., links between the social domain and the item domains.

2. Within-domain links are user-user links in the social domain (social connections) and the item-item links in each item domain.

- *M. Jiang, P. Cui, X. Chen, W. Zhu and S. Yang are with the Department of Computer Science and Technology in Tsinghua University, Beijing 100084.*
- *F. Wang is with Computer Science and Engineering Department in University of Connecticut, 371 Fairfield Way, Unit 4155, Storrs, CT 06269.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

Cross-domain links reflect users' characteristics in different ways. For example, a cross-domain link from a user to a web post about iPhones shows his/her short-term interest in iPhones, and a cross-domain link from him/her to a label "iPhone Fan" implies his/her long-term interest in iPhones. A basic assumption is that the more auxiliary knowledge we have, the more we know about the users, thereby enabling more accurate estimates of tie strength. When a user and his/her friend have many common user labels, we assume a greater tie strength and expect them to be more similar in terms of their web post adoption behaviors. Even if the web post domain is extremely sparse, we may still produce effective recommendations by transferring auxiliary knowledge from other item domains through the social domain.

Thus, knowledge transfer procedures among multiple item domains in social networks should focus on updating tie strength in the social domain, but this is complicated by challenges associated with jointly modeling multiple relational domains, discovering transferable knowledge, and improving recommendations in the target domain.

The following characteristics of the domains considered in this paper are challenging to deal with when developing approaches to recommendation.

(1) The domains are relational. Social network data provide social connections between users, semantic similarity between two items of the same type, and item adoptions by users. The issue of how to represent the user-user links, item-item links, and user-item links poses a challenge to method capability.

(2) The domains are heterogeneous. Heterogeneity is a challenging issue in social recommendation. Within-domain links can be directed ("following" links in the social domain) or undirected (semantic similarity links in the item domains). Cross-domain links can be signed (indicating a positive or negative connotation, such as web-post adoptions and rejections) or unsigned (user-label adoptions). The issue of how to transfer knowledge across heterogeneous domains poses a challenge to method comprehensibility.

(3) The domains are variously sparse. This data sparsity is essentially caused by the large amounts of users and items as well as the time and attention scarcity of these users. It is challenging to try to use relatively dense auxiliary information to help predict sparse links in the target domain.

(4) Items in the domains have varying transferability. Traditional literature often assumes that the most popular items have better transferability. However, later in this work, we will show that this assumption is incorrect. Therefore, transferable knowledge selection approaches for enhancing performance constitutes a literature gap.

To address the above challenges, we propose an innovative *Hybrid Random Walk* (HRW) method for transferring knowledge from auxiliary item domains according to a star-structured configuration to improve social recommendations in a target domain. HRW estimates weights for (1) links between user nodes within the social domain, and (2) links between user nodes in the social domain and item nodes in the item domain. The weights respectively represent (1) tie strength between users and (2) the probability of a user adopting or rejecting an item. Our proposed method integrates knowledge from multiple relational domains and alleviates sparsity and cold-start issues. The key contributions are:

(1) We discover counterintuitive transferability distribution in auxiliary item domains. Besides popularity, we find more meaningful factors, i.e., behavioral consistency with web post adoptions and social connections. These factors have been incorporated into our method.

(2) We propose a novel method to transfer knowledge across multiple relational domains on social networks, incorporating heterogeneous graphs with different types of links. This method can be naturally applied to graph-based applications such as social networks, information networks, and biological networks.

(3) Extensive experimentation on a large real social dataset demonstrates that HRW produces significantly superior recommendations for web posts on social networks. In terms of providing recommendations to cold-start users, only 30% of historical data from the web-post domain is necessary to achieve a comparable performance to that of an approach that makes use of user-label data.

The remainder of this paper is organized as follows. Section 2 discusses related works. Section 3 provides some background and preliminary concepts regarding transferability. Section 4 describes the methodology of the HRW approach with Section 5 providing experimental results. Section 6 concludes.

2 Related Works

In this section, we survey related works and note the literature gap that our research addresses.

2.1 Cross-Domain Collaborative Filtering

Collaborative filtering (CF) techniques are a common approach to recommendation, and have been applied in real recommender systems [11][12][13]. Based on probabilistic matrix factorizations [1], approaches have been proposed to improve recommendation by jointly factorizing a trust network and a user-item matrix [14][2]. One particular contextual model learns both individual preference and interpersonal influence to estimate the probability of item adoption [15]. However, when dealing with information overload, CF does not consider the interplay of users and multiple types of items, so it often suffers from data sparsity and cold start issues.

Recommender systems benefit from new information that goes beyond the user-item matrix [16][17][18][19]. Berkovsky et al. [20] deployed several mediation approaches for importing and aggregating user rating vectors from different domains. Gao et al. [21] conducted

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

recommendation via a cluster-level latent factor model. A joint model of tensor factorization was proposed by Chen et al. to simultaneously recommend users, movies, and tags [22]. However, social applications are different from movie recommenders. Social relation drives information diffusion and adoption [23][24]. Only with the consideration of tie strength can recommender systems better understand users' behavioral intentions. Social Matchbox [25] proposed a latent factor matrix factorization model that treats users' side information (user profiles) and social information as feature vectors to determine user similarity. Facebook uses cross-domain data (user profiles and new feeds) for their recommenders [26]. Sedhain et al. noted that side information is very important [27]. Auxiliary item domains are more complicated than side information. Rich auxiliary user-item interactions, including editing user labels, sharing videos, and joining groups, should be incorporated into a more relational, random walk-based model than factorization-based methods. Fortunately, the advantage of random walk models in utilizing auxiliary information has been proved by empirical research [28].

2.2 Transfer Learning for Recommendation

Adomavicius and Tuzhilin [29] reviewed CF-based, content-based, and hybrid recommendation methods. Their work predicts that auxiliary information will play an important role in the future of recommender systems. Transfer learning provides the key idea of using knowledge from auxiliary domains [30][31][32][33][34][35] and has been used in various ways. Transferring collaborative knowledge from MovieLens can reduce the sparsity problem in recommending movies in Netflix [36]. Book ratings and movie ratings can also be used collaboratively: transferring book ratings can improve movie rating prediction [37]. A recent work by Jing et al. provides a probabilistic collective factorization model to handle sparse data in different settings of knowledge transfer [38]. Transfer learning methods often utilize users' consistent individual preference to bridge two domains by the set of user nodes. However, in social networks, tie strength between users is the key factor utilized to bridge two item domains. We reconsider the representation of social networks, using a hybrid star-structured graph to incorporate within-domain and cross-domain links. Using the social domain as the bridge between the item domains is unique among existing works.

2.3 Random Walk Algorithms and Models

The random walk concept has been widely applied in recommender systems. Random walk based approaches effectively incorporate auxiliary information [28]. Tong et al. [39] proposed a computationally efficient random walk algorithm. ItemRank [40], a random walk-based scoring algorithm, was used to rank products according to expected user preferences. TrustWalker [10] defined and measured the confidence of a recommendation with a random walk model to combine the trust-based and CF approaches. Chen et al. [28] proposed a random walk algorithm to handle both positive and negative comments with the guarantee of convergence. In social networks, social relations and multiple item domains naturally form a star-structured high-order heterogeneous graph [41][42]. In this paper, we develop a random walk-based algorithm on such complex graphs to transfer knowledge from rich, auxiliary domains to a target domain. The biggest difference with respect to previous works is that we use social ties as the fundamental bridge to connect item domains in social networks.

TABLE 1
Dataset Statistics: Amount and Density

Domain	Object	Cross-domain link		Within-domain link
		Accept (+)	Refuse (–)	
user	1,427,214	-	-	20,240,902 9.9×10^{-6}
web post	3,023,609	18,249,207 4.2×10^{-6}	33,608,036 7.8×10^{-6}	- -
user label	5,715	7,604,679 9.3×10^{-4}	- -	- -

3 Preliminaries on Transferability

In this section, we introduce our social dataset of multiple item domains and demonstrate the existence of transferability from auxiliary domains to a target domain.

3.1 Dataset and Distributions

The dataset for this research was crawled in January 2011 from Tencent Weibo (t.qq.com). We crawled data from users who own at least one user label. While the website allows users to have, at most, 10 user labels, the average number of user labels per user was 5.3. The average number of web posts per user was 12.8. We did not filter any social relationships. The average number of friends per user was 14.2.

Table 1 summarizes the data statistics. We used a 5-minute time window to derive negative links. That is, if a user had two adopting behaviors (sharing the web posts) in 5 minutes, we assumed that the user ignored the rest of posts that he/she received in the time window. Thus, besides the two positive user-post links, we noted several negative links. The data indicates that although both web-post and user-label domains are sparse, the latter is denser.

Figure 2 shows (1) distributions of user and post frequency, (2) distributions of user and label frequency, and (3) distributions of follower and followee frequency. We note that the data has smooth distributions, which look like power law relationships in log-log scale. Our dataset has no spiky outliers.

3.2 Symbols and Notations

A real-world example of a second-order hybrid star-structured graph in our dataset is presented in Figure 3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

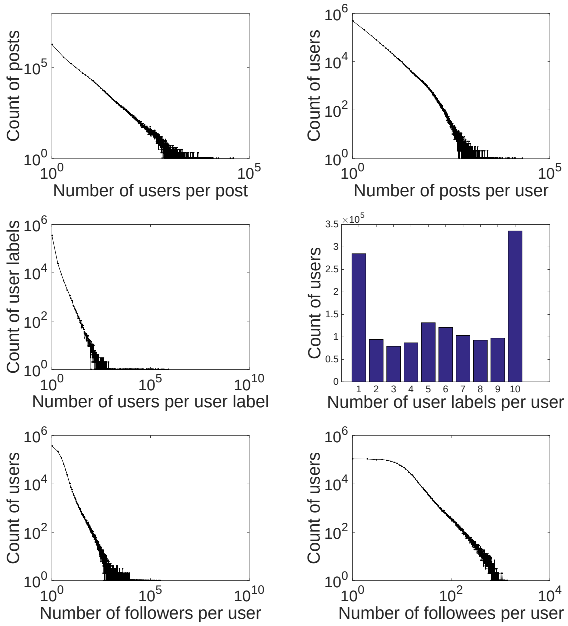


Fig. 2. Smooth power-law-like frequency distributions; our dataset has no spiky outliers. A user often adopts only one or ten (the maximum) user labels.

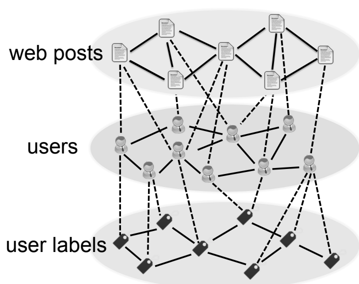


Fig. 3. Our social network data represented by a second-order hybrid star-structured graph, where within-domain links are between users, posts, and labels, and cross-domain links are between user and post, user and label.

It differs from traditional star-structured graphs [41] that do not include entity relationships within each domain. Our hybrid graph considers both within-domain and cross-domain entity relationships.

Table 2 summarizes the symbols used throughout the paper to denote the five subgraphs in Figure 3:

- $\mathcal{G}^{(\mathcal{U})} = \{\mathcal{U}, \mathcal{E}^{(\mathcal{U})}\}$, where $\mathcal{E}^{(\mathcal{U})}$ represents the edge set linking the nodes in $\mathcal{U}$;
- $\mathcal{G}^{(\mathcal{P})} = \{\mathcal{P}, \mathcal{E}^{(\mathcal{P})}\}$, where $\mathcal{E}^{(\mathcal{P})}$ represents the edge set linking the nodes in $\mathcal{P}$;
- $\mathcal{G}^{(\mathcal{T})} = \{\mathcal{T}, \mathcal{E}^{(\mathcal{T})}\}$, where $\mathcal{E}^{(\mathcal{T})}$ represents the edge set linking the nodes in $\mathcal{T}$;
- $\mathcal{G}^{(\mathcal{UP})} = \{\mathcal{U} \bigcup \mathcal{P}, \mathcal{E}^{(\mathcal{UP})}\}$, where $\mathcal{E}^{(\mathcal{UP})}$ represents the edges linking the nodes in $\mathcal{U}$ and $\mathcal{P}$;
- $\mathcal{G}^{(\mathcal{UT})} = \{\mathcal{U} \bigcup \mathcal{T}, \mathcal{E}^{(\mathcal{UT})}\}$, where $\mathcal{E}^{(\mathcal{UT})}$ represents the edges linking the nodes in $\mathcal{U}$ and $\mathcal{T}$.

To conceptualize user relationships in $\mathcal{G}^{(\mathcal{U})}$, consider the relevance from user u_i to u_j as

$$w_{ij}^{(\mathcal{U})} = \begin{cases} 1 & \text{if user } u_i \text{ is a friend of } u_j \text{ or follows } u_j, \\ 0 & \text{otherwise.} \end{cases}$$

TABLE 2
Symbols and Notations

Symbol	Notation
u_i	The i-th user
$\mathcal{U} = \{u_1, u_2, \cdots, u_m\}$	The set of users
p_i	The i-th web post
$\mathcal{P} = \{p_1, p_2, \cdots, p_n\}$	The set of web posts
t_i	The i-th user label
$\mathcal{T} = \{t_1, t_2, \cdots, t_l\}$	The set of user labels
d_{ij}	The j-th item in i-th domain
$\mathcal{D}_i = \{d_{i1}, d_{i2}, \cdots, d_{i\|\mathcal{D}_i\|}\}$	The set of items in i-th domain
$\mathcal{D} = \{D_1, D_2, \cdots, D_N\}$	The set of item domains

To compute web-post relationships in $\mathcal{P}$, we use a Term Frequency-Inverse Document Frequency (TF-IDF) representation vector for each post $\mathrm{b}_i = [b_{i1}, \cdots, b_{ik}, \cdots, b_{iK}]^\top$ in matrix $\mathbf{B}$ (where K is the size of vocabulary), and then measure the semantic similarity between post b_i and b_j as

$$w_{ij}^{(\mathcal{P})} = \frac{\sum_k b_{ik} b_{jk}}{\sqrt{\sum_k {b_{ik}}^2}\sqrt{\sum_k {b_{jk}}^2}}$$

For user labels, we measure relationships using the Jaccard similarity. Assume that labels t_i and t_j appear in c_i and c_j tweets as a word, and co-appear in c_{ij} tweets. Then, the semantic relationship is computed as

$$w_{ij}^{(\mathcal{T})} = \frac{c_{ij}}{c_i + c_j - c_{ij}}$$

Thus, we have constructed three similarity matrices $\mathbf{W}^{(\mathcal{U})} = \{w_{ij}^{(\mathcal{U})}\}$, $\mathbf{W}^{(\mathcal{P})} = \{w_{ij}^{(\mathcal{P})}\}$ and $\mathbf{W}^{(\mathcal{T})} = \{w_{ij}^{(\mathcal{T})}\}$ to encode edge weights for three within-domain subgraphs.

Further, we have two cross-domain subgraphs $\mathcal{G}^{(\mathcal{UP})}$ and $\mathcal{G}^{(\mathcal{UT})}$, whose edge weights need to be estimated. Since web posts can be adopted or rejected but user labels are edited by users, both positive and negative user-post links exist, but only positive user-label links exist. These links are presented as undirected edges $e_{ij}^{(\mathcal{UP})}$ and $e_{ij}^{(\mathcal{UT})}$. Their weights are determined as follows.

$$\begin{aligned} w_{ij}^{(\mathcal{UP})^+} &= \begin{cases} 1 & \text{if user } u_i \text{ adopts web post } \rho_j, \\ 0 & \text{otherwise;} \end{cases} \\ w_{ij}^{(\mathcal{UP})^-} &= \begin{cases} 1 & \text{if user } u_i \text{ rejects web post } \rho_j, \\ 0 & \text{otherwise;} \end{cases} \\ w_{ij}^{(\mathcal{UT})^+} &= \begin{cases} 1 & \text{if user } u_i \text{ adopts web post } t_j, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, we obtain the three weight matrices $\mathbf{W}^{(\mathcal{UP})^+}=\{w_{ij}^{(\mathcal{UP})^+}\}$, $\mathbf{W}^{(\mathcal{UP})^-}=\{w_{ij}^{(\mathcal{UP})^-}\}$, and $\mathbf{W}^{(\mathcal{UT})^+}=\{w_{ij}^{(\mathcal{UT})^+}\}$.

3.3 Transferability of User Labels

Here, we conduct data analysis to demonstrate that (1) the auxiliary, user-label domain can be transferred to predict a target web-post domain; that is, user-label interactions are in some degree consistent with user-post interactions; (2) user-label interactions are also consistent

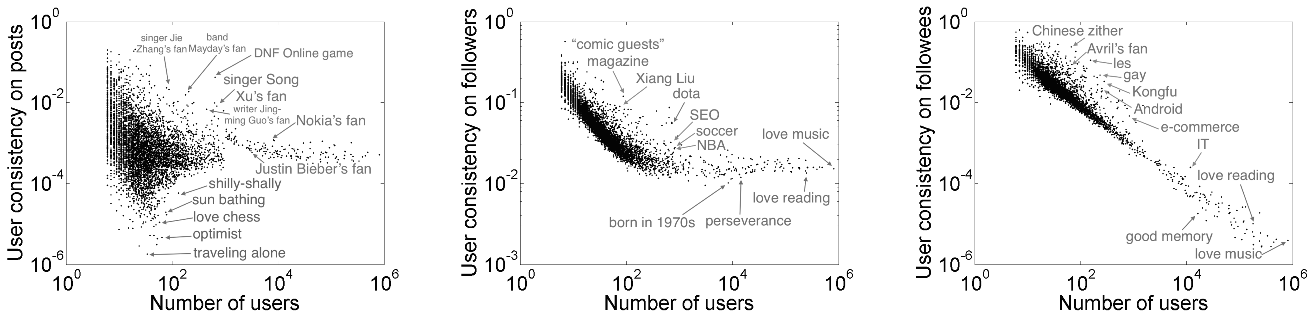


Fig. 4. Popular user labels do not transfer the most knowledge: different user labels have different consistency in terms of adopting similar posts, sharing followers, and having the same followees. Surprisingly, the consistency is often in reverse ratio with the popularity. Thus, selecting transferable user labels is important.

with user-user interactions in the social domain; and (3) not every label can be transferred, and the most popular labels are not the most transferable.

First, we define the popularity of a user label j as the number of users who adopt the label as follows:

$$popularity(j) = \sum_i w_{ij}^{(\mathcal{UT})^+}$$

Second, we define three kinds of consistency for the user label j: (1) consistency with users' post contents, i.e., the average similarity of the web posts between a pair of users who have the label j, denoted by

$$cons_{post}(j) = \frac{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} \frac{\mathbf{w}_{i,:}^{(\mathcal{UP})^+} \mathbf{w}_{k,:}^{(\mathcal{UP})^+}}{||\mathbf{w}_{i,:}^{(\mathcal{UP})^+}|| ||\mathbf{w}_{k,:}^{(\mathcal{UP})^+}||}}{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} 1}$$

(2) consistency with users' followers, i.e., the similarity of followers between a pair of users, denoted by

$$cons_{follower}(j) = \frac{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} \frac{\mathbf{w}_{:,i}^{(\mathcal{U})} \mathbf{w}_{:,k}^{(\mathcal{U})}}{||\mathbf{w}_{:,i}^{(\mathcal{U})}|| ||\mathbf{w}_{:,k}^{(\mathcal{U})}||}}{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} 1}$$

(3) consistency with users' followees, i.e., the similarity of followees between a pair of users, denoted by

$$cons_{followee}(j) = \frac{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} \frac{\mathbf{w}_{i,:}^{(\mathcal{U})} \mathbf{w}_{k,:}^{(\mathcal{U})}}{||\mathbf{w}_{i,:}^{(\mathcal{U})}|| ||\mathbf{w}_{k,:}^{(\mathcal{U})}||}}{\sum_{i \neq k; w_{ij}^{(\mathcal{UT})^+}, w_{kj}^{(\mathcal{UT})^+} > 0} 1}$$

Figure 4 plots user labels comparing their *popularity* and three kinds of *consistency*. Surprisingly, the consistency score is often in reverse ratio with the popularity. To be more specific, users' behaviors with respect to tagging themselves with labels such as "Nokia's fan" and "Justin Bieber's fan" are consistent with their behaviors on adopting web posts, while their behaviors with respect to tagging themselves with psychological characteristics, such as "traveling alone" and "optimist", are inconsistent. Users with labels such as "soccer", "NBA", and "SEO" usually have similar followers since they tend to be active users, who often post messages related to these labels. Meanwhile, users with labels such as "IT" and "e-commerce" have similar followees since they often connect to accounts that are famous in the related areas. Note that the most popular user labels such as "love music" and "love reading" are not consistent with user behaviors in the target domain and social domain. These labels cannot generally enrich the knowledge in users' behaviors due to poor transferability. Thus, we see that user labels can be transferred due to their consistency with other behaviors of adopting posts and sharing friends, and we find that selecting transferable user labels is important for knowledge transfer.

4 Hybrid Random Walk Algorithm

In this section, we introduce our random walk-based method on social recommendation. Owing to data sparsity in the target domain, traditional Bipartite Random Walk (BRW) algorithms cannot accurately derive user tie strength to predict user behaviors in the target domain [43] [44]. Fortunately, we have auxiliary domains in which user ties are formed for the same reason as in the target domain: homophily, trust, and influence. The key idea is to utilize rich knowledge from the auxiliary domains to better describe user tie strength and then more precisely predict user behaviors. Thus, we derive HRW algorithms on star-structured graphs.

4.1 On Hybrid Second-Order Star-Structured Graph

We derive a random walk algorithm to predict missing links on $\mathcal{G}^{(\mathcal{UP})}$ and $\mathcal{G}^{(\mathcal{UT})}$, which includes both within-domain and cross-domain random walks. For $\mathcal{G}^{(\mathcal{U})}$, $\mathcal{G}^{(\mathcal{P})}$ and $\mathcal{G}^{(\mathcal{T})}$, we derive steady-state distributions [39], indicating the intrinsic relevance among users, posts and labels. For a standard random walk model, a walker starts from the i-th vertex and iteratively jumps to other vertices with transition probabilities $\mathbf{p}_i = \{p_{i1}, \cdots, p_{in}\}$. After reaching the steady state, the probability of the walker staying at the j-th vertex corresponds to the relevance score of vertex j to i. Specifically, the transition probability matrices are computed as the row-normalized weight matrices:

$$\begin{aligned}
\mathbf{P}^{(\mathcal{U})} &= (\mathbf{D}^{(\mathcal{U})})^{-1}\mathbf{W}^{(\mathcal{U})} \\
\mathbf{P}^{(\mathcal{P})} &= (\mathbf{D}^{(\mathcal{P})})^{-1}\mathbf{W}^{(\mathcal{P})} \\
\mathbf{P}^{(\mathcal{T})} &= (\mathbf{D}^{(\mathcal{T})})^{-1}\mathbf{W}^{(\mathcal{T})}
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

where we denote the degree matrices of cross-domain links by $\mathbf{D}^{(\mathcal{UP})^+}$, $\mathbf{D}^{(\mathcal{UP})^-}$, and $\mathbf{D}^{(\mathcal{UT})^+}$.

The final steady-state probability matrices can be obtained by iterating the following updates:

$$\begin{aligned}\mathbf{R}^{(\mathcal{U})}(t+1) &= \alpha\mathbf{P}^{(\mathcal{U})}\mathbf{R}^{(\mathcal{U})}(t)+(1-\alpha)\mathbf{I}\\ \mathbf{R}^{(\mathcal{P})}(t+1) &= \beta\mathbf{P}^{(\mathcal{P})}\mathbf{R}^{(\mathcal{P})}(t)+(1-\beta)\mathbf{I}\\ \mathbf{R}^{(\mathcal{T})}(t+1) &= \gamma\mathbf{P}^{(\mathcal{T})}\mathbf{R}^{(\mathcal{T})}(t)+(1-\gamma)\mathbf{I}\end{aligned}$$

where $\mathbf{R}^{(\mathcal{U})}(t)$, $\mathbf{R}^{(\mathcal{P})}(t)$, $\mathbf{R}^{(\mathcal{T})}(t)$, $\mathbf{R}^{(\mathcal{U})}(t+1)$, $\mathbf{R}^{(\mathcal{P})}(t+1)$ and $\mathbf{R}^{(\mathcal{T})}(t+1)$ are the state probability matrices at time t and $t+1$; and $0 \le \alpha, \beta, \gamma \le 1$ are the prior probabilities that the random walker will leave its current state. It can be easily shown that the above iterations will finally converge to the steady state matrices when $t \to \infty$.

$$\begin{aligned}\mathbf{R}^{(\mathcal{U})} &= (1-\alpha)(\mathbf{I}-\alpha\mathbf{P}^{(\mathcal{U})})^{-1}\\ \mathbf{R}^{(\mathcal{P})} &= (1-\beta)(\mathbf{I}-\beta\mathbf{P}^{(\mathcal{P})})^{-1}\\ \mathbf{R}^{(\mathcal{T})} &= (1-\gamma)(\mathbf{I}-\gamma\mathbf{P}^{(\mathcal{T})})^{-1}\end{aligned}$$

For cross-domain links, we compute the transition probability matrices as

$$\begin{aligned}\mathbf{P}^{(\mathcal{UP})^+} &= (\mathbf{D}^{(\mathcal{UP})^+})^{-1}\mathbf{W}^{(\mathcal{UP})^+}\\ \mathbf{P}^{(\mathcal{UP})^-} &= (\mathbf{D}^{(\mathcal{UP})^-})^{-1}\mathbf{W}^{(\mathcal{UP})^-}\\ \mathbf{P}^{(\mathcal{UT})^+} &= (\mathbf{D}^{(\mathcal{UT})^+})^{-1}\mathbf{W}^{(\mathcal{UT})^+}\end{aligned}$$

where elements $p_{ij}^{(\mathcal{UP})^+}$ and $p_{ij}^{(\mathcal{UP})^-}$ represent the transition probability that user u_i will adopt/ignore post p_j; and $p_{ij}^{(\mathcal{UT})^+}$ represents the transition probability that user u_i will adopt label t_j. Now, we simultaneously determine relevance scores $\mathbf{R}^{(\mathcal{U})} = \{r_{ij}^{(\mathcal{U})}\}$, between each pair of users, which finally reflects the tie strength on the real user graph. Element $r_{ij}^{(\mathcal{U})}$ represents the probability that a random walker jumps from user u_i to u_j. Now, we consider the above transition paths and estimate the transition probabilities $p_{ij}^{(\mathcal{UP})^+}$, $p_{ij}^{(\mathcal{UP})^-}$, $p_{ij}^{(\mathcal{UT})^+}$, and $r_{ij}^{(\mathcal{U})}$ of one step random walk over $\mathcal{G}^{(\mathcal{UP})}$, $\mathcal{G}^{(\mathcal{UT})}$, and $\mathcal{G}^{(\mathcal{U})}$ as

$$\begin{aligned}p_{ij}^{(\mathcal{UP})^+} &= \delta\sum_{u_k\in\mathcal{U}} r_{ik}^{(\mathcal{U})}p_{kj}^{(\mathcal{UP})^+} + (1-\delta)\sum_{p_k\in\mathcal{P}} p_{ik}^{(\mathcal{UP})^+}r_{kj}^{(\mathcal{P})}\\ p_{ij}^{(\mathcal{UP})^-} &= \delta\sum_{u_k\in\mathcal{U}} r_{ik}^{(\mathcal{U})}p_{kj}^{(\mathcal{UP})^-} + (1-\delta)\sum_{p_k\in\mathcal{P}} p_{ik}^{(\mathcal{UP})^-}r_{kj}^{(\mathcal{P})}\\ p_{ij}^{(\mathcal{UT})^+} &= \eta\sum_{u_k\in\mathcal{U}} r_{ik}^{(\mathcal{U})}p_{kj}^{(\mathcal{UT})^+} + (1-\eta)\sum_{t_k\in\mathcal{T}} p_{ik}^{(\mathcal{UT})^+}r_{kj}^{(\mathcal{T})}\\ r_{ij}^{(\mathcal{U})} &= \tau^{(\mathcal{P})}(\mu\sum_{p_k\in\mathcal{P}} p_{ik}^{(\mathcal{UP})^+}p_{jk}^{(\mathcal{UP})^+}\\ &\quad +(1-\mu)\sum_{p_k\in\mathcal{P}} p_{ik}^{(\mathcal{UP})^-}p_{jk}^{(\mathcal{UP})^-})\\ &\quad +\tau^{(\mathcal{T})}\sum_{t_k\in\mathcal{T}} p_{ik}^{(\mathcal{UT})^+}p_{jk}^{(\mathcal{UT})^+} + \tau^{(\mathcal{U})}\sum_{u_k\in\mathcal{U}} r_{ik}^{(\mathcal{U})}r_{kj}^{(\mathcal{U})}\end{aligned}$$

where $0 \le \delta, \eta, \mu, \tau^{(\mathcal{P})}, \tau^{(\mathcal{T})}, \tau^{(\mathcal{U})} \le 1$ are the parameters for trading off different transition routes. Note that for the update of cross-domain transition probability matrices, we consider the two types of routes shown in Figure 5. We also assume that the update of the cross-domain probability matrices will affect the within-domain probability matrix of the user subgraph. The updating rules the within-domain probability matrices consider the three routes shown in Figure 6.

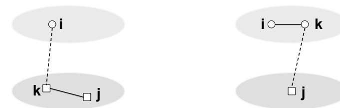


Fig. 5. Transition routes we consider when updating a cross-domain transition probability matrix.

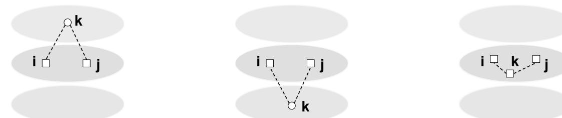


Fig. 6. Transition routes when updating a within-domain transition probability matrix on social relation.

Our proposed model assumes that the update of cross-domain transition probability affects only the within-domain transition probabilities of the user graph, because the user tie strength is influenced by (1) common posts, (2) common labels, and (3) social relationships. We claim that the cross-domain links (adoption behaviors) do not affect the within-domain transition probability of other item domains. The rationale is that the transition probability of items (posts, labels, videos, etc.) should be derived by their semantic similarity, which would not be changed by the users who adopt them. Therefore, the HRW method updates cross-domain links between the user and all types of items as well as within-domain links between users (not items).

We can further give matrix formulations for the update of transition probability from time t to $t+1$.

$$\begin{aligned}&\mathbf{P}^{(\mathcal{UP})^+}(t+1) = \delta\mathbf{R}^{(\mathcal{U})}(t)\mathbf{P}^{(\mathcal{UP})^+}(t) + (1-\delta)\mathbf{P}^{(\mathcal{UP})^+}(t)\mathbf{R}^{(\mathcal{P})}\\ &\mathbf{P}^{(\mathcal{UP})^-}(t+1) = \delta\mathbf{R}^{(\mathcal{U})}(t)\mathbf{P}^{(\mathcal{UP})^-}(t) + (1-\delta)\mathbf{P}^{(\mathcal{UP})^-}(t)\mathbf{R}^{(\mathcal{P})}\\ &\mathbf{P}^{(\mathcal{UT})^+}(t+1) = \eta\mathbf{R}^{(\mathcal{U})}(t)\mathbf{P}^{(\mathcal{UT})^+}(t) + (1-\eta)\mathbf{P}^{(\mathcal{UT})^+}(t)\mathbf{R}^{(\mathcal{T})}\\ &\mathbf{R}^{(\mathcal{U})}(t+1) = \tau^{(\mathcal{P})}(\mu\mathbf{P}^{(\mathcal{UP})^+}(t)\mathbf{P}^{(\mathcal{UP})^+}(t)^T\\ &\quad +(1-\mu)\mathbf{P}^{(\mathcal{UP})^-}(t)\mathbf{P}^{(\mathcal{UP})^-}(t)^T)\\ &\quad +\tau^{(\mathcal{T})}\mathbf{P}^{(\mathcal{UT})^+}(t)\mathbf{P}^{(\mathcal{UT})^+}(t)^T + \tau^{(\mathcal{U})}\mathbf{R}^{(\mathcal{U})}(t)\mathbf{R}^{(\mathcal{U})}(t)^T\end{aligned}$$

With graphs $\mathcal{G}^{(\mathcal{U})}$, $\mathcal{G}^{(\mathcal{UP})}$, and $\mathcal{G}^{(\mathcal{UT})}$, the corresponding transition matrices $\mathbf{R}^{(\mathcal{U})}$, $\mathbf{P}^{(\mathcal{UP})^+}$, $\mathbf{P}^{(\mathcal{UP})^-}$, and $\mathbf{P}^{(\mathcal{UT})^+}$ are computed for the next random walk step. Algorithm 1 summarizes the procedure of the second-order star-structured graph-based iteratively random walk method for predicting post and label adoptions. The space complexity of this algorithm is $O(m^2+n^2+l^2+2m(n+l))$, and the time complexity is $O((m^2+4m(n+l)+2(n^2+l^2))mT)$, where m, n, and l are the number of users, posts, and labels, respectively, and T is the number of iterations. Since the matrices are usually sparse and $m, n \gg l$, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

space complexity is $O((m+n)^2)$ and the time complexity is $O((m+n)ET)$, where E is the number of links between users and posts.

Algorithm 1 Iterative Adoption Prediction through Random Walk over a Second-Order Star-Structured Graph

Require: $0 \leq \alpha, \beta, \gamma, \delta, \eta, \mu, \tau^{(\mathcal{P})}, \tau^{(\mathcal{T})}, \tau^{(\mathcal{U})} \leq 1$
1: Construct graphs $\mathcal{G}^{(\mathcal{U})}$, $\mathcal{G}^{(\mathcal{P})}$, $\mathcal{G}^{(\mathcal{T})}$, $\mathcal{G}^{(\mathcal{UP})}$, $\mathcal{G}^{(\mathcal{UT})}$
2: Compute transition probabilities $\mathbf{P}^{(\mathcal{U})}$, $\mathbf{P}^{(\mathcal{P})}$ and $\mathbf{P}^{(\mathcal{T})}$
3: Derive steady-state $\mathbf{R}^{(\mathcal{U})}$, $\mathbf{R}^{(\mathcal{P})}$ and $\mathbf{R}^{(\mathcal{T})}$
4: Initialize transition probability matrices $\mathbf{P}^{(\mathcal{UP})^+}(0)$, $\mathbf{P}^{(\mathcal{UP})^-}(0)$ and $\mathbf{P}^{(\mathcal{UT})^+}(0)$
5: **for** $t = 1 : T$ **do**
6: Compute tie strength matrix $\mathbf{R}^{(\mathcal{U})}(t)$ and transition probability matrices $\mathbf{P}^{(\mathcal{UP})^+}(t)$, $\mathbf{P}^{(\mathcal{UP})^-}(t)$ and $\mathbf{P}^{(\mathcal{UT})^+}(t)$
7: **end for**
8: **Output:** Final transition probability matrices $\mathbf{R}^{(\mathcal{U})}$, $\mathbf{P}^{(\mathcal{UP})^+}$, $\mathbf{P}^{(\mathcal{UP})^-}$ and $\mathbf{P}^{(\mathcal{UT})^+}$

As discussed in Section 3.3, items in auxiliary domains have different degrees of transferability. Here, we adopt a widely-used vote-counting scheme introduced in 1770 by Jean Charles de Borda to select the transferable items for improving the transfer learning framework. We call this variant of the HRW method *HRW-Borda*.

In particular, each feature, such as popularity, consistency with users' post contents, consistency with users' followers, and consistency with users' followees, provides a ranked voting system for user labels. The number of points given to the label candidates for each ranking is determined by the number of candidates standing in the ranked list. The total number of points from all the systems evaluates the transferability of the user label. We select the labels with the largest 1,000 values of the transferability and only use the partial user-label links.

4.2 On Hybrid High-Order Star-Structured Graph

In previous discussions, we assumed that two types of item domains, web post and user label, are associated with each user. However, online social networks are an unprecedented comprehensive platform with a number of different types of User-Generated Content (UGC), e.g., posts, labels, music, and movies. Figure 1 shows a typical example of a hybrid high-order star-structured graph with four different types of UGC. In this case, the second-order graph is insufficient for describing all the UGC. Our random walk strategy can be easily extended to higher-order cases. We represent the following subgraphs contained in the high-order hybrid graph.

- $\mathcal{G}^{(\mathcal{U})} = \{\mathcal{U}, \mathcal{E}^{(\mathcal{U})}\}$, where $\mathcal{E}^{(\mathcal{U})}$ represents the edge set linking the nodes in $\mathcal{U}$
- $\mathcal{G}^{(\mathcal{D}_i)} = \{\mathcal{D}_i, \mathcal{E}^{(\mathcal{D}_i)}\}$, where $\mathcal{E}^{(\mathcal{D}_i)}$ represents the edge set linking the nodes in $\mathcal{D}_i$, $i = 1, \cdots, N$
- $\mathcal{G}^{(\mathcal{UD}_i)} = \{\mathcal{U} \bigcup \mathcal{D}_i, \mathcal{E}^{(\mathcal{UD}_i)}\}$, where $\mathcal{E}^{(\mathcal{UD}_i)}$ represents the edges linking the nodes in $\mathcal{U}$ and $\mathcal{D}_i$, $i = 1, \cdots, N$

With respect to $\mathcal{G}^{(\mathcal{U})}$ and $\{\mathcal{G}^{(\mathcal{D}_i)}\}_{i=1}^N$, we construct their corresponding edge weight matrices $\mathbf{W}^{(\mathcal{U})}$ and $\{\mathbf{W}^{(\mathcal{D}_i)}\}_{i=1}^N$. Thus, the within-domain transition probability matrices can be obtained by ($i = 1, \cdots, N$)

$$\begin{aligned} \mathbf{P}^{(\mathcal{U})} &= (\mathbf{D}^{(\mathcal{U})})^{-1}\mathbf{W}^{(\mathcal{U})} \\ \mathbf{P}^{(\mathcal{D}_i)} &= (\mathbf{D}^{(\mathcal{D}_i)})^{-1}\mathbf{W}^{(\mathcal{D}_i)} \end{aligned}$$

where $\mathbf{D}^{(\mathcal{U})}$ and $\{\mathbf{D}^{(\mathcal{D}_i)}\}_{i=1}^N$ are the degree matrices induced by $\mathbf{W}^{(\mathcal{U})}$ and $\{\mathbf{W}^{(\mathcal{D}_i)}\}_{i=1}^N$. The final steady-state probability matrices can be iteratively calculated by

$$\begin{aligned} \mathbf{R}^{(\mathcal{U})}(t+1) &= \alpha\mathbf{P}^{(\mathcal{U})}\mathbf{R}^{(\mathcal{U})}(t) + (1-\alpha)\mathbf{I} \\ \mathbf{R}^{(\mathcal{D}_i)}(t+1) &= \beta_i\mathbf{P}^{(\mathcal{D}_i)}\mathbf{R}^{(\mathcal{D}_i)}(t) + (1-\beta_i)\mathbf{I} \end{aligned}$$

where $i = 1, 2, \cdots, N,\ 0 \leq \alpha, \beta_1, \cdots, \beta_N \leq 1$.

For the cross-domain subgraphs $\{\mathcal{G}^{(\mathcal{UD}_i)}\}_{i=1}^N$, we compute the edge weight matrices $\{\mathbf{W}^{(\mathcal{UD}_i)}\}_{i=1}^N$ based on the user interactions with other item domains $\{\mathcal{D}_i\}_{i=1}^N$. Thus, the cross-domain transition probability matrices can be computed as

$$\begin{aligned} \mathbf{P}^{(\mathcal{UD}_i)^+} &= \left(\mathbf{D}^{(\mathcal{UD}_i)^+}\right)^{-1}\mathbf{W}^{(\mathcal{UD}_i)^+} \\ \mathbf{P}^{(\mathcal{UD}_i)^-} &= \left(\mathbf{D}^{(\mathcal{UD}_i)^-}\right)^{-1}\mathbf{W}^{(\mathcal{UD}_i)^-} \end{aligned}$$

where $i = 1, \cdots, N$. When updating the cross-domain transition probability matrices, we consider the transition routes shown in Figures 5 and 6, so that they can be updated using the following rules:

$$\begin{aligned} \mathbf{P}^{(\mathcal{UD}_i)^+}(t+1) &= \delta_i\mathbf{R}^{(\mathcal{U})}(t)\mathbf{P}^{(\mathcal{UD}_i)^+}(t) \\ &= +(1-\delta_i)\mathbf{P}^{(\mathcal{UD}_i)^+}(t)\mathbf{R}^{(\mathcal{D}_i)} \\ \mathbf{P}^{(\mathcal{UD}_i)^-}(t+1) &= \delta_i\mathbf{R}^{(\mathcal{U})}(t)\mathbf{P}^{(\mathcal{UD}_i)^-}(t) \\ &= +(1-\delta_i)\mathbf{P}^{(\mathcal{UD}_i)^-}(t)\mathbf{R}^{(\mathcal{D}_i)} \\ \mathbf{R}^{(\mathcal{U})}(t+1) &= \sum_{\mathcal{D}_i \in \mathcal{D}} \tau_i\mu_i\mathbf{P}^{(\mathcal{UD}_i)^+}(t)\mathbf{P}^{(\mathcal{UD}_i)^+}(t)^T \\ &\quad + \sum_{\mathcal{D}_i \in \mathcal{D}} \tau_i(1-\mu_i)\mathbf{P}^{(\mathcal{UD}_i)^-}(t)\mathbf{P}^{(\mathcal{UD}_i)^-}(t)^T \\ &\quad + \tau^{(\mathcal{U})}\mathbf{R}^{(\mathcal{U})}(t)\mathbf{R}^{(\mathcal{U})}(t)^T \end{aligned}$$

where $0 \leq \delta_i, \mu_i, \tau_i \leq 1$ are the trade-off parameters and $i = 1, 2, \cdots, N$. For a domain $\mathcal{D}_i$ without negative user-item links, we set $\mu_i = 1$ to update $\mathbf{R}^{(\mathcal{U})}$.

Algorithm 2 summarizes the procedure of a random walk on a high-order hybrid star-structured graph for predicting user adoptions on different item domains. The space complexity of this algorithm is $O(m^2 + 2m\sum|\mathcal{D}_i| + \sum|\mathcal{D}_i|^2)$ and the time complexity is $O((m^2 + 4m\sum|\mathcal{D}_i| + 2\sum|\mathcal{D}_i|^2)mT)$, where T is the number of iterations.

5 Experiments

In this section, we give the experimental results of applying our HRW methods to our dataset, which is a social network with two item domains, web posts, and user labels. We evaluate the performance on the (1) social recommendation problem, i.e., predicting positive

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

Algorithm 2 Random Walk over a High-Order Hybrid Star-Structured Graph

Require: $0 \le \alpha, \{\beta_i\}_{i=1}^N, \{\delta_i\}_{i=1}^N, \{\mu_i\}_{i=1}^N, \{\tau_i\}_{i=1}^N \le 1$

Construct $\mathcal{G}^{(\mathcal{U})}$, $\{\mathcal{G}^{(\mathcal{D}_i)}\}_{i=1}^N$, $\{\mathcal{G}^{(\mathcal{U}\mathcal{D}_i)}\}_{i=1}^N$

Compute $\mathbf{P}^{(\mathcal{U})}$ and $\{\mathbf{P}^{(\mathcal{D}_i)}\}_{i=1}^N$

Derive $\mathbf{R}^{(\mathcal{U})}$ and $\{\mathbf{R}^{(\mathcal{D}_i)}\}_{i=1}^N$

Initialize $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^+}(0)\}_{i=1}^N$ and $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^-}(0)\}_{i=1}^N$

for $t = 1 : T$ **do**

  Compute $\mathbf{R}^{(\mathcal{U})}(t)$, $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^+}(t)\}_{i=1}^N$ and $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^-}(t)\}_{i=1}^N$

end for

Output: Final transition probability matrices $\mathbf{R}^{(\mathcal{U})}$, $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^+}\}_{i=1}^N$ and $\{\mathbf{P}^{(\mathcal{U}\mathcal{D}_i)^-}\}_{i=1}^N$

TABLE 3
Variants of Our HRW and Existing BRW Methods

Algorithm	$\mathcal{G}^{(\mathcal{U})}$	$\mathcal{G}^{(\mathcal{P})}$	$\mathcal{G}^{(\mathcal{T})}$
HRW-Borda	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, combined
HRW-Popular	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, *popularity*
HRW-Cons-Post	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, $cons_{post}$
HRW-Cons-Follower	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, $cons_{follower}$
HRW-Cons-Followee	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, $cons_{followee}$
HRW-All	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	$\mathbf{W}^{(\mathcal{T})}$, all
BRW-R_U-P (TrustWalker)	$\mathbf{R}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	×
BRW-R_U	$\mathbf{R}^{(\mathcal{U})}$	×	×
BRW-W_U-P	$\mathbf{W}^{(\mathcal{U})}$	$\mathbf{W}^{(\mathcal{P})}$	×
BRW-W_U (ItemRank)	$\mathbf{W}^{(\mathcal{U})}$	×	×
BRW-P	×	$\mathbf{W}^{(\mathcal{P})}$	×

and negative user-post links, and the (2) user cold-start problem, i.e., some users are new and the number of training links is zero. The results demonstrate the effectiveness of our approach.

5.1 Experimental Settings

Here, we present our experimental settings regarding performance evaluation metrics, parameter settings, and comparative algorithms.

5.1.1 Evaluation Metrics

We adopted three evaluation measures: reconstruction error, prediction accuracy, and ranking-based metrics.

- Error-based metrics: Mean Absolute Error (MAE) defined as
$$\frac{1}{N}\sum_{u_i,p_j}(|p_{ij}^{(\mathcal{UP})^+} - \hat{p}_{ij}^{(\mathcal{UP})^+}| + |p_{ij}^{(\mathcal{UP})^-} - \hat{p}_{ij}^{(\mathcal{UP})^-}|)$$
and Root Mean Square Error (RMSE) defined as
$$(\frac{1}{N}\sum_{u_i,p_j}(|p_{ij}^{(\mathcal{UP})^+} - \hat{p}_{ij}^{(\mathcal{UP})^+}|^2 + |p_{ij}^{(\mathcal{UP})^-} - \hat{p}_{ij}^{(\mathcal{UP})^-}|^2))^{\frac{1}{2}},$$
where $p_{ij}^{(\mathcal{UP})^+}$ and $p_{ij}^{(\mathcal{UP})^-}$ are the ground truth adoption and rejection of user u_i on item p_j in the testing set; $\hat{p}_{ij}^{(\mathcal{UP})^+}$ and $\hat{p}_{ij}^{(\mathcal{UP})^-}$ denote the prediction results, and N denotes the size of the testing set. All p values are ground truth values equal to either 0 or 1; $\hat{p}$ values are probability scores in the range $[0,1]$. The MAE and RMSE evaluate the error between the prediction and ground-truth values.
- Accuracy-based metrics: Precision, recall, and F1 refer to the harmonic mean of precision and recall:
$$precision = \frac{|\{(u_i,p_j)|\hat{p}_{ij}^{(\mathcal{UP})^+} > \hat{p}_{ij}^{(\mathcal{UP})^-}, p_{ij}^{(\mathcal{UP})^+} = 1\}|}{|\{(u_i,p_j)|\hat{p}_{ij}^{(\mathcal{UP})^+} > \hat{p}_{ij}^{(\mathcal{UP})^-}\}|}$$
$$recall = \frac{|\{(u_i,p_j)|\hat{p}_{ij}^{(\mathcal{UP})^+} > \hat{p}_{ij}^{(\mathcal{UP})^-}, p_{ij}^{(\mathcal{UP})^+} = 1\}|}{|\{(u_i,p_j)|p_{ij}^{(\mathcal{UP})^+} = 1\}|}$$
$$F1 = \frac{2 \times precision \times recall}{precision + recall},$$
where $p_{ij}^{(\mathcal{UP})^+}$ and $p_{ij}^{(\mathcal{UP})^-}$ are the ground truth adoption and rejection of user u_i on item p_j in the testing set, and $\hat{p}_{ij}^{(\mathcal{UP})^+}$ and $\hat{p}_{ij}^{(\mathcal{UP})^-}$ denote the prediction result.
- Ranking-based metrics: Mean Average Precision (MAP)@K. This involves, for the top K recommended items, evaluating the mean of average precision. In our experiments, we set K as 1, 3, 5, 10 and 20.

5.1.2 Parameter Settings

With regard to parameter settings, we have δ and η as the relative weights of user tie strength over item similarity on user-post and user-label links, μ as the relative weight of positive samples over negative samples on user-post links, and $\tau^{(\mathcal{P})}$ and $\tau^{(\mathcal{T})}$ as the relative weights of cross-domain links from web-post domain over user-label domain on influencing user tie strength. All parameters range from 0 to 1, which we tune according to greedy search. We find the best 6 parameters δ, η, μ, $\tau^{(\mathcal{U})}$, $\tau^{(\mathcal{P})}$, and $\tau^{(\mathcal{T})}$ to reduce the error metrics. In our parameter settings, we find that when $\tau^{(\mathcal{U})}$ is as small as 0.05, our HRW achieves the best performance, which proves that most user behaviors are influenced by direct friends or some indirect friends. Meanwhile, the best settings of $\tau^{(\mathcal{P})}$ and $\tau^{(\mathcal{T})}$ that represent the relative weights of user tie strength from item similarity on the post and label domain are in the middle of the range.

5.2 Comparative Algorithms

We want to answer the following two questions through our experiments: (1) Does our method, which transfers knowledge from auxiliary domains, work better in social recommendation and cold start scenarios than methods that do not transfer this knowledge? (2) Does selecting transferable items in the auxiliary domain work better than using all the items or the most popular items? Therefore, as shown in Table 3, we have two series of baseline methods to prove the effectiveness of our HRW methods. The methods use within-domain links in the social domain, web-post domain, and user-label domain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

from $\mathbf{W}^{(\mathcal{U})}$, $\mathbf{W}^{(\mathcal{P})}$, and $\mathbf{W}^{(\mathcal{T})}$, and learn user tie strength by updating $\mathbf{R}^{(\mathcal{U})}$, i.e., weights on social links.

To answer the first question, we compare with baseline methods of a BRW that predict user-post links without rich knowledge from the user-label domain. We also carefully implement a state-of-the-art matrix factorization method. The algorithms tested were:

- BRW-R_U-P (TrustWalker [10]), which predicts user-post links, with post similarity utilized and user tie strength updated.
- BRW-R_U, which predicts links, with tie strength updated on a bipartite graph but no post similarity.
- BRW-W_U-P, which predicts links, with user relation and post similarity but no tie strength.
- BRW-W_U (ItemRank [40]), which uses only social relations to compute the weights on user-post links.
- BRW-P, which learns only post similarity to predict user-post links with a random walk model.
- TLLSM, a very recent method named Transfer Learning with Latent Space Matching [18], which incorporates rich user and item information into recommendation, adding a "matching-based" regularization term on factorization:

$$\begin{aligned}\min \quad & ||\mathbf{W} \odot (\mathbf{P}^{(\mathcal{UP})^+} - \mathbf{UV})||^2 \\ & + ||\mathbf{W} \odot (\mathbf{P}^{(\mathcal{UT})^+} - \mathbf{XY})||^2 \\ & + \lambda(||\mathbf{U}||^2 + ||\mathbf{V}||^2 + ||\mathbf{X}||^2 + ||\mathbf{Y}||^2) \\ & + \beta \sum_m arctan(||\mathbf{U}(m,:) - \mathbf{X}(m,:)||^2)\end{aligned}$$

where $\mathbf{W}$ is the indicator, λ and β are regularization parameters; $\mathbf{U}$ and $\mathbf{X}$ represent user preference vectors; and $\mathbf{V}$ and $\mathbf{Y}$ represents item latent vectors.

For the second question, we have the following HRW methods that select transferable labels.

- HRW-Borda, which uses Borda count to vote for the l_{top} best labels in the following four ranking lists made by features.
- HRW-Popular, which uses labels with the l_{top} highest values of *popularity*.
- HRW-Cons-Post, which uses labels with the l_{top} highest values of *consistency with users' posts*.
- HRW-Cons-Follower, which uses labels with the l_{top} highest values of *consistency with users' followers*.
- HRW-Cons-Followee, which uses labels with the l_{top} highest values of *consistency with users' followees*.
- HRW-All, which uses all user-label links.

5.3 Experimental Results

In this section, we first compare the performance of our proposed HRW method and other comparable algorithms on predicting missing user-post links for social recommendation. Second, we discuss the transferability of different items in the auxiliary domain and the performance of our item selection method. Third, we discuss the transferability across domains, i.e., (1) how user tie strength works as a bridge, (2) how transferring auxiliary information performs, and (3) how positive/negative samples help. Finally, we show how our method solves the cold-start problem and sheds light on the usefulness of auxiliary information.

5.3.1 *Social Recommendation Performance*

Here, we demonstrate the method's performance when predicting missing links in social scenarios. We conduct hold-out experiments by randomly selecting 80% of user-post links for training and the remaining for testing, while user-label links are completely utilized. This random selection is carried out 20 times independently, and we report the average and variance values.

Comparisons on accuracy: Table 4 compares the performances of our method with its different configurations and with the above baselines, respectively, on the average results and standard deviations of evaluations including the MAE, RMSE, precision, recall, and F1 scores. Note that (1) our method outperforms existing approaches in experimental trials and is insensitive to initialization; and (2) HRW-Borda, which uses Borda count to select auxiliary items, performs best among all methods. Specifically, compared with BRW, we observe that:

- BRW-W_U reduces MAE by 18.4% compared with BRW-P, an item-based recommendation implemented by the random walk algorithm. BRW-W_U exploits user dependent preferences from friendships and performs better than the collaborative approach on large social datasets. BRW-R_U reduces MAE by 10.5% compared with BRW-W_U, which updates user tie strength on the social domain with user-post links. BRW-R_U-P reduces MAE by 14.3% compared with BRW-W_U, which learns both within-domain links (post similarities) and cross-domain links (user-post links) to update user tie strength. This is because the motivations of user behavior on social networks are that (1) users like to adopt web posts that highly correlate with those adopted before, and (2) users like to adopt posts recommended by their friends or followers with high tie strength. BRW-R_U-P combines these two aspects to solve the social recommendation problem.
- HRW-All reduces MAE by 22.1% compared with BRW-R_U-P. This is consistent with our assumption that in social networks, user tie strength is shaped by multiple relational domains such as web-post and user-label domains. Our method effectively utilizes auxiliary information to formulate the weighted user graph, and performs better than BRW in solving the sparsity problem of user-post link prediction.
- HRW outperforms the most recent matrix factorization based approach TLLSM with side information of posts and users (i.e., word distributions of posts and user-label links) incorporated, reducing MAE by 27.9%. User behaviors on social networks stem from the interrelationships among users, tight or

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

TABLE 4
Results of Our Methods and Baselines on Social Recommendation (Predicting Missing User-Post Links)

Algorithm	MAE	RMSE	Precision	Recall	F1 measure
HRW-Borda	**0.195±1.3e-3**	**0.226±2.6e-3**	**0.912±4.7e-3**	**0.771±3.5e-3**	**0.791±5.1e-4**
HRW-Popular	0.278±3.7e-3	0.306±1.8e-3	0.866±6.7e-4	0.728±9.1e-4	0.729±2.8e-3
HRW-Cons-Post	0.215±3.2e-3	0.249±3.5e-3	0.900±4.7e-3	0.758±2.5e-3	0.774±3.8e-3
HRW-Cons-Follower	0.250±4.6e-3	0.285±3.7e-3	0.881±1.9e-3	0.741±9.5e-4	0.751±3.6e-3
HRW-Cons-Followee	0.227±2.2e-3	0.254±2.5e-3	0.891±8.1e-4	0.752±1.1e-3	0.764±3.0e-3
HRW-All	0.260±3.5e-3	0.296±3.4e-3	0.874±4.4e-3	0.738±2.8e-3	0.741±5.7e-4
BRW-R_U-P	0.334±3.1e-3	0.357±1.5e-3	0.832±6.3e-4	0.699±4.1e-3	0.689±9.9e-4
BRW-R_U	0.349±3.5e-3	0.371±1.5e-3	0.831±8.8e-4	0.696±1.5e-3	0.682±4.6e-3
BRW-W_U-P	0.377±3.6e-3	0.403±3.9e-3	0.813±3.4e-3	0.677±1.7e-3	0.656±3.7e-3
BRW-W_U	0.390±3.9e-3	0.419±4.0e-3	0.802±3.7e-3	0.668±4.4e-3	0.643±2.9e-3
BRW-P	0.478±3.5e-3	0.499±4.1e-3	0.754±6.3e-4	0.629±4.4e-3	0.583±3.6e-3
TLLSM	0.361±2.6e-3	0.385±1.7e-3	0.816±2.7e-3	0.685±4.0e-3	0.668±2.6e-3

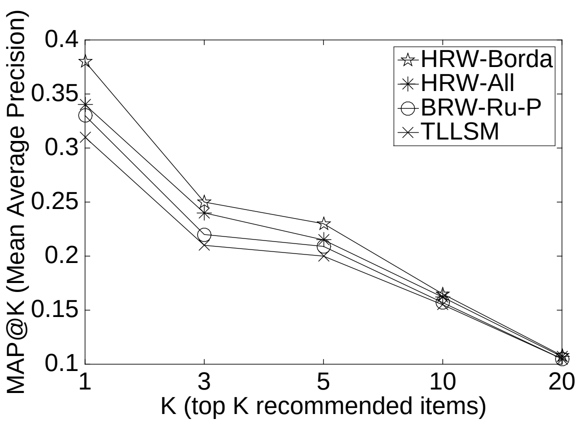


Fig. 7. The results of MAP@K for different K for HRW-Borda, HRW-all, BRW-R_U-P and TLLSM. The figure shows that HRW methods have better MAP scores; HRW-Borda outperforms any competitor. When K is smaller, the improvement on MAP@K increases.

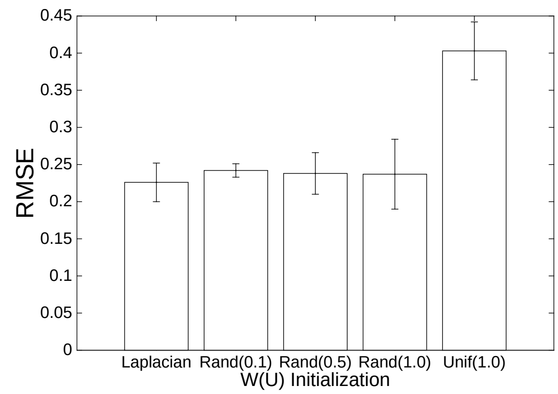


Fig. 8. HRW method is insensitive to tie strength initializations with RMSE remaining reasonably consistent across different HRW-Borda configurations as follows. The left four bars are HRW-Borda with different configurations of $\mathbf{W}^{(U)}$: the first uses the Laplacian matrix of the binary graph; the other three randomize user-user weights in a given range. The rightmost bar represents a configuration of fixed 1 for every user tie (=BRW-W_U-P).

loose, that have been naturally shaped. However, TLLSM fails to consider user tie strength.

Comparisons on ranking metrics: Figure 7 shows MAP evaluations of our HRW-Borda and HRW-All methods and baselines for recommending the top K items. We observe that (1) our HRW methods produce higher MAP scores than the best BRW method (BRW-R_U-P) and the matrix factorization method TLLSM, and (2) the HRW-Borda outperforms all the competitors. Though we note that the improvement on MAP@20 is tiny when K is as big as 20, our HRW methods improve MAP@1 by 16.9% and MAP@3 by 17.4%. These results demonstrate that HRW methods significantly improve the item ranking performance in social recommendation scenarios.

Comparisons on insensitivity: Figure 8 compares the RMSE of HRW-Borda method, with multiple settings of tie strength matrix $\mathbf{W}^{(U)}$ initialization, including (1) Laplacian, i.e., the Laplacian matrix of degree matrix of the social graph; (2) Rand(x), $x \in \{0.1, 0.5, 1.0\}$, i.e., initializing non-zero entries in $\mathbf{W}^{(U)}$ with random values from 0 to x; and (3) Unif(1.0), i.e., initializing all entries in $\mathbf{W}^{(U)}$ with fixed values 1. We observe that

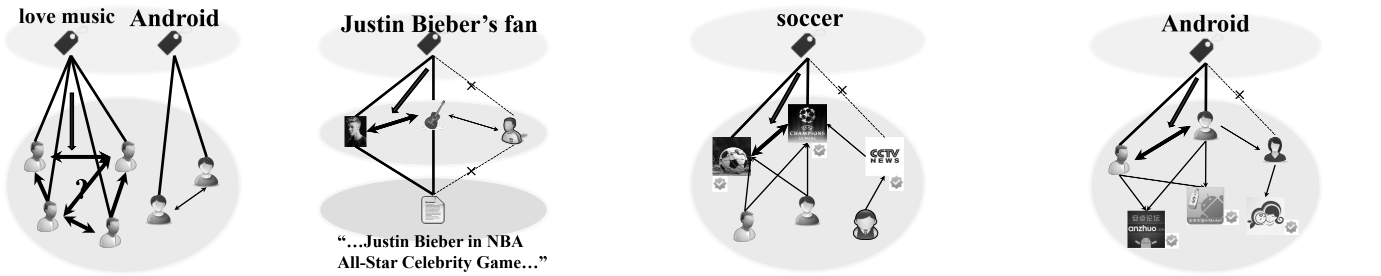
- The first four bars show similar RMSE values, indicating that HRW method is *insensitive* to tie strength initializations. Moreover, the error bars show that when x is bigger, the variance (error) of RMSE is bigger. Our default setting, Laplacian, gives small RMSE and small variance.
- The last bar represents the all-1 configuration, which gives much larger RMSE. This representation is equivalent to the BRW-R_U-P method, showing that neglecting social information decreases the prediction accuracy.

These results demonstrate that incorporating social relations significantly improves the recommendation performance and that the performance of HRW is insensitive to the tie strength initializations.

5.3.2 Performance on Auxiliary Item Transferability

In addition to the comparison with BRW methods, we compare different configurations of HRW algorithms on selecting transferable items in the auxiliary domain.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering



(a) Popularity (b) Consistency with posts (c) Consistency with followers (d) Consistency with followees

Fig. 9. What user labels can reflect real tie strengths? Users with popular common labels do not have to be strongly connected. However, the consistency of labels with posts reflects users' common preferences, and the consistency of labels with follower/followee sets reflects users' social communities. These factors are illustrated in the above figure. A combined strategy is applied in our method.

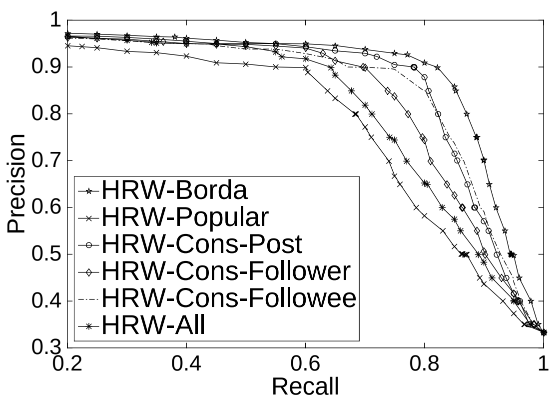


Fig. 10. Precision vs. recall for several HRW methods. HRW-Borda gives better precision-recall results than other HRW methods. Considering all the features, including popularity and consistency, HRW-Borda improves the prediction performance by selecting highly transferable items in the auxiliary domain.

Table 4 shows that the MAE values of HRW-Cons-Post, HRW-Cons-Follower, and HRW-Cons-Followee are 22.6%, 10.0%, and 18.3% smaller, respectively, than the MAE of HRW-Popular. Further, in Figure 10, we plot precision-recall curves for all the algorithms and find that HRW-Borda reaches the best, almost perfect values. All the above 5 methods select the same number of auxiliary items (user labels). HRW-Popular assumes that users with the same, more popular labels are often more tightly connected. However, in Figure 9(a), while "love music" is adopted by 815,166 users and "Android" is adopted by 10,653 users, the social relations between users with "Android" are much denser than those between users with "love music". Although "love music" is popular, it is ordinary in terms of appearing in users' label sets and weak for reflecting users' tastes on making social friends and adopting posts. We further give case studies to discuss how to select user labels with high transferability in auxiliary domains.

- HRW-Cons-Post uses the consistency with users' posts feature. It selects the labels based on post similarity. For example, in Figure 9(b), users labeled "Justin Bieber's fan" retweet messages about Justin's performance in NBA All-Star Celebrity Game. HRW-Cons-Post assumes that they have stronger tie strength because of their similar preferences on post content.
- HRW-Cons-Follower uses the consistency with users' followers feature. It selects the labels based on the possession of common followers. In Figure 9(c), the famous accounts with label "soccer" like Tencent Soccer and CCTV5 UEFA have 13,522 common followers (soccer fans), who are strongly connected and often interact. Although both CCTV News and CCTV5 UEFA are hosted by CCTV, they have no common label and they do not often interact.
- HRW-Cons-Followee uses the consistency with users' followees feature. It selects the labels based on the possession of common followees. For example, in Figure 9(d), users with "Android" connect to a few big accounts such as Android Forum and Android Market. They are tightly connected and often retweet messages about new Apps from famous accounts. HRW-Cons-Followee has better performance than HRW-Cons-Follower, because it gives weights to user relations among a large population of ordinary users, while HRW-Cons-Follower can only reflect tie strength between famous users.

In our final version of the HRW method, HRW-Borda combines all the above features of user labels and uses the standard Borda count to select the most transferable user labels. Compared with HRW-All, which uses all the items in the auxiliary domain, HRW-Borda reduces the MAE by 25.0%. Further, compared with the best results from BRW methods and previous approaches, our HRW-Borda reduces MAE by as much as 41.6%.

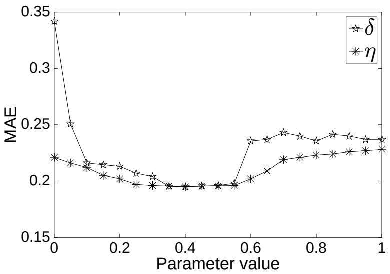
5.3.3 *Performance on Domain Transferability*

Here, we use the experimental results to answer the three questions below related to the transferability across different domains.

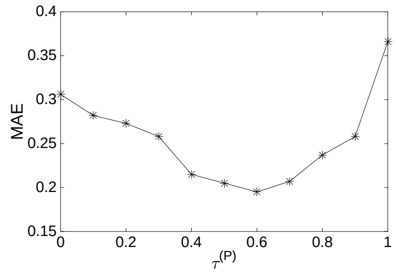
(1) Are item similarity and tie strength important in predicting user-post links?

On the web-post domain, δ is the weight of the tie strength over web-post similarity. If δ increases, users

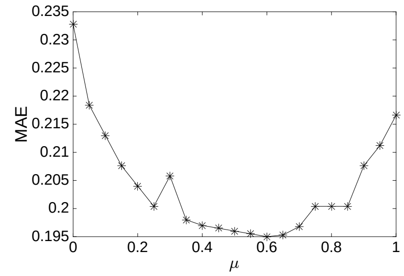
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering



(a) MAE v.s. δ and η



(b) MAE v.s. $\tau^{(\mathcal{P})}$



(c) MAE v.s. μ

Fig. 11. Charts of MAE against parameters. These charts provide the following insights: (a) Tune weights of tie strength and item similarity on cross-domain link prediction. (b) Tune weights of knowledge from web-post and user-label domains transferred to tie strength. (c) Tune weights of positive and negative user-post samples on tie strength.

are more likely to accept the recommended posts for their social relationships than for their preferences. η is the corresponding weight on the user-label domain. Figure 11(a) shows the performances when varying δ and η from 0 to 1. For δ, $\delta = 0$ means the prediction without considering item similarity, while $\delta = 1$ means the prediction without considering tie strength. From the figure we can clearly observe a valley when δ is around 0.4, which means that incorporating both tie strength and post similarity can significantly improve performance.

(2) Are user-post and user-label links important in influencing user tie strength?

$\tau^{(\mathcal{P})}$ and $\tau^{(\mathcal{T})}$ correspond to the weights of knowledge learned from user-post and user-label links on calculating the tie strength, with $\tau^{(\mathcal{P})} + \tau^{(\mathcal{T})} = 1$. Figure 11(b) shows the MAE when varying $\tau^{(\mathcal{P})}$ from 0 to 1. When $\tau^{(\mathcal{P})} = 1$, i.e., when we discard the knowledge transferred from user-label links on tie strength, MAE is around 0.37. When $\tau^{(\mathcal{P})} = 0$, i.e., $\tau^{(\mathcal{T})} = 1$, MAE is around 0.31, which may indicate that the user-label domain is more helpful in predicting tie strength since it is easier for friendly users to share the same user labels than the same posts. The minimum MAE ($\tau^{(\mathcal{P})} = 0.6$) suggests that knowledge from user-label links provides potential clues on user-post link prediction through user tie strength; thus, recommender systems should incorporate these two kinds of links.

(3) Are negative samples of web-post domain helpful?

μ is the relative weight of positive samples from web-post domain that influences user tie strength, while $1-\mu$ is the relative weight of negative samples. Figure 11(c) shows the changing curve of MAE when varying μ from 0 to 1. When $\mu = 1$ (only train positive samples) and $\mu = 0$ (only train negative samples), the MAE is higher than when taking both samples for training ($\mu = 0.6$). This means that considering both positive and negative samples in the recommendation is helpful for algorithm performance. MAE is lower when we have only negative samples than when we have only positive ones.

The above discussion proves that our method is reasonable and effective, which considers comprehensive factors on user behavior and user tie strength, while considering both positive and negative samples.

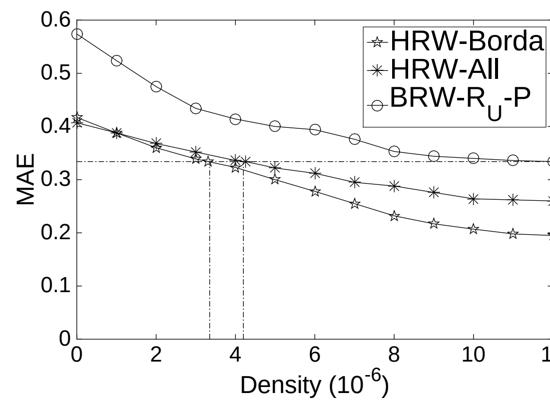


Fig. 12. MAE vs. density for HRW-Borda, HRW-all, and BRW-R_U-P. HRW-Borda needs only 27.6% of the historical data and HRW-All needs only 35.5% of the historical data in the target domain to reach the same performance as BRW-R_U-P, which does not transfer knowledge from user-label links. Our HRW gives promising insights to support the solving of the cold-start problem.

5.3.4 Performance on User Cold Start Problem

In this section, we conduct experiments to test the performance of three methods on recommending items for cold-start users: (1) BRW-R_U-P, which has performed the best in predicting missing links among all BRW methods; (2) HRW-All, which uses all auxiliary user-label links to help predict users' behaviors in the target domain; and (3) HRW-Borda, which selects highly transferable items from the auxiliary domain to transfer strong and proper knowledge. We control the density of training entries of testing users (the percentage of training user-post links per testing user, testing item). The experimental results are shown in Figure 12. If no training entry is hidden, where the density of training entries of testing users is 1.2×10^{-5}, HRW-Borda reduces MAE by 41.6% compared with the baseline (0.195 over 0.334). If we hide all training entries, i.e., the density of training entries of testing users is zero, which means the testing users are new in the application, without previous behaviors in the historical data, our method still reduces MAE by 27.4% (0.417 over 0.574). Furthermore, from Figure 12, we observe that:

- Our transfer learning methods HRW-Borda and HRW-All need only 27.6% and 35.5% of the training entries (3.31×10^{-6} and 4.26×10^{-6} dense) to reach the same level performance of BRW-R_U-P with the whole training set (1.2×10^{-5} dense). With user labels of a new user, our method needs only 3-day historical data to reach the same recommendation performance of 10-day data without labels. Therefore, if we motivate new users to add several user labels, the transferred knowledge from the user-label domain would greatly improve user experience on personal recommendation services.
- When the training data set is empty, HRW-All performs a bit better than HRW-Borda, mainly because it uses more rich information to evaluate users' tie strength. With the number of user-label links increasing, the consistency of user labels becomes important, and thus, HRW-Borda, which selects transferable labels, performs much better than HRW.

6 Conclusion

In this paper, we addressed the problems of data sparsity and cold start in social recommendation. We reconsidered the problem from the transfer learning perspective and alleviated the data sparsity problem in a target domain by transferring knowledge from other auxiliary social relational domains. By considering the special structures of multiple relational domains in social networks, we proposed an innovative HRW method on a star-structured graph, which is a general method to incorporate complex and heterogeneous link structures.

We conducted extensive experiments on a large real-world social network dataset and showed that the proposed method greatly boosts the social recommendation performance. In particular, we gained improvement in web-post recommendation by transferring knowledge from the user-label domain for the user tie strength updating process, compared with the recommendation methods, which only use information from the web-post domain. In addition, we demonstrated that, by using only 27.6% of the available information in the target domain, our method achieves comparable performance with methods that use all available information in the target domain without transfer learning. The proposed method and insightful experiments indicate a promising and general way to solve the data sparsity problem.

References

[1] Y. Koren, R. Bell, and C. Volinsky, ”Matrix factorization techniques for recommender systems,” in Computer, vol. 42, pp. 30–37, 2009.
[2] H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King, ”Recommender system with social regularization,” in WSDM, pp. 287–296, 2011.
[3] I. Konstas, V. Stathopoulos, and J. M. Jose, ”On social networks and collaborative recommendation,” in SIGIR, pp. 195–202, 2009.
[4] J. Wang, H. Zeng, Z. Chen, H. Lu, L. Tao, and W. Ma, ”ReCoM: Reinforcement clustering of multi-type interrelated data objects,” in SIGIR, pp. 274–281, 2003.
[5] J. Brown, A. J. Broderick, and N. Lee, ”Word of mouths communication within online communities: conceptualizing the online social network,” in Journal of Interactive Marketing, vol. 21, pp. 2–20, 2007.
[6] J. Leskovec, A. Singh, and J. Kleinberg, ”Patterns of influence in a recommendation network,” in LNCS, vol. 3918, pp. 380–389, 2006.
[7] X. Yang, H. Steck, and Y. Liu. ”Circle-based recommendation in online social networks.” in SIGKDD, pp. 1267–1275, 2012.
[8] X. Qian, H. Feng, G. Zhao, and T. Mei. ”Personalized Recommendation Combining User Interest and Social Circle.” in IEEE TKDE, vol. 26, no. 7, pp. 1763–1777, 2014.
[9] P. Massa and P. Avesani, ”Trust-aware recommender systems,” in RecSys, pp. 17–24, 2007.
[10] M. Jamali and M. Ester, ”TrustWalker: a random walk model for combining trust-based and item-based recommendation,” in SIGKDD, pp. 397–406, 2009.
[11] N. N. Liu, M. Zhao, and Q. Yang, ”Probabilistic latent preference analysis for collaborative filtering,” in CIKM, pp. 759–766, 2009.
[12] M. Harvey, M. J. Carman, I. Ruthven, and F. Crestanig, ”Bayesian latent variable models for collaborative item rating prediction,” in CIKM, pp. 699–708, 2011.
[13] J. Noel, S. Sanner, K. N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad, and E. D. Penna. ”New objective functions for social collaborative filtering.” in WWW, pp. 859–868, 2012.
[14] H. Ma, H. Yang, M.R. Lyu, and I. King, ”SoRec: social recommendation using probabilistic matrix factorization,” in CIKM, pp. 931–940, 2008.
[15] M. Jiang, P. Cui, R. Liu, F. Wang, W. Zhu, and S. Yang, ”Scalable Recommendation with Social Contextual Information,” in IEEE TKDE, vol. 26, no. 11, pp. 2789–2802, 2014.
[16] P. Winoto and T. Tang. ”If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations,” in New Generation Computing, vol. 26, no. 3, pp. 209–225, 2008.
[17] J. Tang, S. Wu, J. Sun, and H. Su. ”Cross-domain collaboration recommendation,” in SIGKDD, pp. 1285–1293, 2012.
[18] C. Li, and S. Lin. ”Matching users and items across domains to improve the recommendation quality,” in SIGKDD, pp. 801-810, 2014.
[19] Y. Shi, M. Larson, and A. Hanjalic. ”Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges,” in ACM Computing Surveys, 47(1):3, 2014.
[20] S. Berkovsky, T. Kuflik, and F. Ricci. ”Cross-domain mediation in collaborative filtering,” in User Modeling, pp. 355–359, 2007.
[21] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo. ”Cross-domain recommendation via cluster-level latent factor model,” in Machine Learning and Knowledge Discovery in Databases, pp. 161-176, 2013.
[22] W. Chen, W. Hsu, and M. L. Lee. ”Making recommendations from multiple domains,” in SIGKDD, pp. 892-900, 2013.
[23] J. Tang, X. Hu, H. Gao, and H. Liu. ”Exploiting local and global social context for recommendation,” in IJCAI, pp. 2712–2718, 2013.
[24] J. Tang, X. Hu, and H. Liu. ”Social recommendation: a review,” in Social Network Analysis and Mining, vol. 3, no. 4, pp. 1113–1133, 2013.
[25] S. Sedhain, S. Sanner, L. Xie, R. Kidd, K. N. Tran, and P. Christen. ”Social affinity filtering: Recommendation through fine-grained analysis of user interactions and activities.” in COSN, pp. 51–62, 2013.
[26] B. Shapira, L. Rokach, and S. Freilikhman. ”Facebook single and cross domain data for recommendation systems,” in User Modeling and User-Adapted Interaction, 23(2-3), pp. 211–247, 2013.
[27] S. Sedhain, S. Sanner, D. Braziunas, L. Xie, and J. Christensen. ”Social collaborative filtering for cold-start recommendations.” in RecSys, pp. 345-348, 2014.
[28] Y. Chen, Y. Lin, Y. Shen, and S. Lin. ”A modified random walk framework for handling negative ratings and generating explanations.” in TIST, vol. 4, no. 1, 2013.
[29] G. Adomavicius and A. Tuzhilin. ”Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” in IEEE TKDE, vol. 17, no. 6, pp. 734–749, 2005.
[30] S. J. Pan and Q. Yang. ”A survey on transfer learning,” in IEEE TKDE, vol. 22, no. 10, pp. 1345-1359, 2010.
[31] B. Cao, N. N. Liu, and Q. Yang. ”Transfer learning for collective link prediction in multiple heterogeneous domains,” in ICML, pp. 159–166. 2010.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2015.2432811, IEEE Transactions on Knowledge and Data Engineering

[32] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G.-R. Xue, Y. Yu, and Q. Yang. "Heterogeneous transfer learning for image classification," in AAAI, 2011.

[33] O. Moreno, B. Shapira, L. Rokach, and G. Shani. "Talmud: transfer learning for multiple domains," in CIKM, pp. 425-434, 2012.

[34] Z. Lu, W. Pan, E. W. Xiang, Q. Yang, L. Zhao, and E. Zhong. "Selective transfer learning for cross domain recommendation," in SDM, pp. 641-649, 2013.

[35] B. Tan, E. Zhong, M. K. Ng, and Q. Yang. "Mixed-transfer: transfer learning over mixed graphs," in SDM, 2014.

[36] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in AAAI, pp. 230–235, 2010.

[37] B. Li, Q. Yang, and X. Xue. "Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction," in IJCAI, vol. 9, pp. 2052–2057, 2009.

[38] H. Jing, A.-C. Liang, S.-D. Lin, and Y. Tsao. "A transfer probabilistic collective factorization model to handle sparse data in collaborative filtering." in ICDM, 2014.

[39] H. Tong, C. Faloutsos, and J. Pan, "Fast random walk with restart and its applications," in ICDM, pp. 613–622, 2006.

[40] M. Gori and A. Pucci, "ItemRank: a random-walk based scoring algorithm for recommender engines," in IJCAI, pp. 2766–2771, 2007.

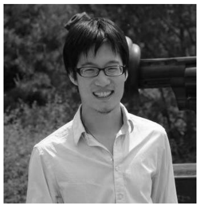
[41] B. Gao, T. Liu, X. Zheng, Q. Cheng, and W. Ma, "Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering," in SIGKDD, pp. 41–50, 2005.

[42] B. Gao, T. Liu, and W. Ma, "Star-structured high-order heterogeneous data co-clustering based on consistent information theory," in ICDM, pp. 880–884, 2006.

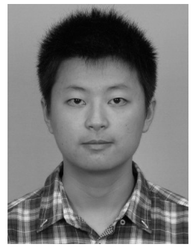
[43] C. Avin, and B. Krishnamachari. "The power of choice in random walks: an empirical study." in Computer Networks 52, no. 1, pp. 44-60, 2008.

[44] I. Safro, P. D. Hovland, J. Shin, and M. M. Strout. "Improving Random Walk Performance." in CSC, pp. 108–112, 2009.

Meng Jiang received the B.E. degree from the Department of Computer Science and Technology of Tsinghua University in 2010. He is a final-year Ph.D. candidate in the Department of Computer Science and Technology of Tsinghua University. His main research interests include data mining, behavior analysis and social network analysis. He visited Carnegie Mellon University from 2012 to 2013. He has published papers on social recommendation, behavioral modeling and suspicious behavior detection in top conferences/journals of the relevant field such as IEEE TKDE, ACM TKDD, ACM SIGKDD, CIKM, PAKDD, etc. He got the best paper finalist in ACM SIGKDD 2014.



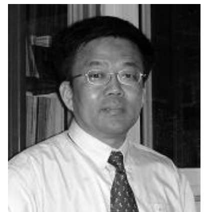
Peng Cui received the Ph.D. degree in computer science in 2010 from Tsinghua University and he is an Assistant Professor at Tsinghua. He has vast research interests in data mining, multimedia processing, and social network analysis. Until now, he has published more than 20 papers in conferences such as SIGIR, AAAI, ICDM, etc. and journals such as IEEE TMM, IEEE TIP, DMKD, etc. Now his research is sponsored by National Science Foundation of China, Samsung, Tencent, etc. He also serves as Guest Editor, Co-Chair, PC member, and Reviewer of several high-level international conferences, workshops, and journals.



Xumin Chen is expected to get his B.E. degree from Department of Computer Science and Technology in Tsinghua University in July 2015. He has won several medals in international programming competitions. He serves as the training coach of National Olympiad in Informatics.

Fei Wang is now Associate Professor in Department of Computer Science and Engineering, University of Connecticut. Before his current position, Dr. Wang was a Research Staff Member in healthcare analytics research group, IBM T. J. Watson Research Center. Dr. Wang got his Ph. D. degree from Department of Automation, Tsinghua University in 2008. After that, he spent one year in School of Computing and Information Science, Florida International University as a postdoc and another year in Department of Statistical Science, Cornell University as a postdoc. His research interests include data mining, machine learning and their applications in healthcare data analytics. He has published over 100 papers on the leading conferences like SIGKDD, SIGIR, ICML, IJCAI, AAAI, SDM, ICDM. He also serves as a referee for many distinguished journals including IEEE TPAMI, IEEE TKDE, DMKD, ACM TKDD and program committee member for many international conferences including KDD, ICDM and SDM. He got the best paper honorable mention in ICDM 2010, best paper finalist in SDM 2011, and finalist for Marco Romani best paper award in AMIA Translational Bioinformatics Summit 2014.

Wenwu Zhu received the Ph.D. degree from Polytechnic Institute of New York University in 1996. He is now a Professor at Tsinghua University. His research interest is wireless/Internet multimedia communication and computing. He worked at Bell Labs during 1996 to 1999. He was with Microsoft Research Asia's Internet Media Group and Wireless and Networking Group as research manager from 1999 to 2004. He was the director and chief scientist at Intel Communication Technology Lab, China. He was also a senior researcher at the Internet Media Group at Microsoft Research Asia. He has published more than 200 referred papers and led 40 patents. He participated in the IETF ROHC WG on robust TCP/IP header compression over wireless links and IEEE 802.16m WG standardization. He currently serves as Chairman of IEEE Circuits and System Society Beijing Chapter, advisory board on the International Journal of Handheld Computing Research, and President of ACM Beijing.



Shiqiang Yang received the B.E. and M.E. degrees from the Department of Computer Science and Technology of Tsinghua University in 1977 and 1983, respectively. He is now a Professor at Tsinghua University. His research interests include multimedia technology and systems, video compression and streaming, content-based retrieval for multimedia information. multimedia content security, and digital right management. He has published more than 100 papers and MPEG standard proposals. Prof. Yang has organized many conferences as program Chair or TPC member including PCM05, MMM06, ICME06, MMSP05, ASWC06, etc.