Item Recommendation for Emerging Online Businesses

Chun-Ta Lu[†], Sihong Xie[†], Weixiang Shao[†], Lifang He[‡] and Philip S. Yu[†§]
[†]Department of Computer Science, University of Illinois at Chicago, IL, USA
[‡]Institute for Computer Vision, Shenzhen University, China
[§]Institute for Data Science, Tsinghua University, Beijing, China.
{clu29,sxie6,wshao4}@uic.edu, lifanghescut@gmail.com, psyu@uic.edu

Abstract

Nowadays, a large number of new online businesses emerge rapidly. For these emerging businesses, existing recommendation models usually suffer from the data-sparsity. In this paper, we introduce a novel similarity measure, *AmpSim* (Augmented Meta Path-based Similarity) that takes both the linkage structures and the augmented link attributes into account. By traversing between heterogeneous networks through overlapping entities, AmpSim can easily gather side information from other networks and capture the rich similarity semantics between entities. We further incorporate the similarity information captured by AmpSim in a collective matrix factorization model such that the transferred knowledge can be iteratively propagated across networks to fit the emerging business. Extensive experiments conducted on real-world datasets demonstrate that our method significantly outperforms other state-of-the-art recommendation models in addressing item recommendation for emerging businesses.

1 Introduction

Collaborative filtering (CF) methods [Koren, 2008; Rong *et al.*, 2014] based on historical user-item interactions have been proven to be one of the most successful solutions for recommendation in developed online businesses. However, it is a different story when employing CF methods for emerging online businesses. Generally speaking, emerging businesses can be online services that are still in its embryonic stage; or mature ones that start to branch into new geographic areas or new categories [Zhang and Yu, 2015]. In an emerging business, available user data are usually too sparse for effective recommendation. Furthermore, how to make accurate prediction for new users with extremely few records still remains a challenge, which is called the "cold start" problem.

Fortunately, different online businesses may share some common users and items. Figure 1 shows an example of an emerging e-commerce site and a developed review site. Note that among four customers in the e-commerce site, two of them also have user accounts in the review site. Besides, two products are shared by both sites. The recommendation task
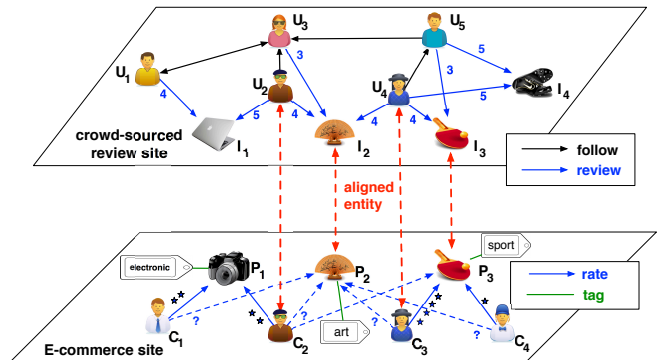

Figure 1: Example of paritally aligned heterogeneous networks, where some entities in one network belong to the same identities in the other network.

in this paper is to predict the rating of a given item by a target customer, as shown in question marks at the bottom of Figure 1. Although user data are extremely sparse in the e-commerce site, abundant knowledge in the more developed review site can be utilized to help recommendation.

Recently, researchers have applied *transfer learning* to CF methods for alleviating the data sparsity in recommender systems [Singh and Gordon, 2008; Luo *et al.*, 2014]. Most transfer learning models in CF assume that there are correspondences between users or items across domains. For example, collective matrix factorization (CMF) [Singh and Gordon, 2008] finds joint low-rank representations by simultaneously factorizing several matrices, sharing parameters among latent factors when an entity participates in multiple relations. However, the prerequisite of entity correspondence across different domains is often hard to satisfy for all entities in real-world scenarios. Manually identifying the entity correspondences is expensive and time-consuming as users may use different names, or an item may be named differently in different online businesses. Different identification algorithms [Kong *et al.*, 2013; Zafarani and Liu, 2013; Lu *et al.*, 2014; Zhang *et al.*, 2015] are designed to automatically mapping entities across domains but they are usually much less accurate. Further, the overlapping users can be biased to more active users. The others can only benefit from the transferred knowledge through the sparse user-item interactions in the target domain, leading to sub-optimal results.

One possible way to handle the above issue is to find similar entities and encode them with similar representations [Wang and Mahadevan, 2011; Long *et al.*, 2014]. However, most existing similarity measures are defined within a single business, and thus they cannot escape from the data-sparsity issue in the emerging business. Moreover, discarding important information between entities, i.e., relationship attributes, may cause degraded performance. How can we define an accurate similarity measure by keeping relationship attributes and further leveraging side information from multiple businesses?

In this paper we introduce the concept of augmented meta path (AMP), which is a sequence of relations between entities and the relations are augmented with link attributes. Taking both the linkage structures and the augmented link attributes into account, we define a novel similarity measure called *AmpSim* that can judiciously capture the rich similarity semantics between entities via AMPs. By traversing between networks through overlapping entities, AmpSim can easily gather side information from other networks to help measure entity similarity. For example, the similarity between customers C_2 and C_3 in Figure 1 can be measured through the AMP instance $C_2 \leftrightarrow U_2 \xrightarrow{\text{[score=4]}} I_2 \xleftarrow{\text{[score=4]}} U_4 \leftrightarrow C_3$, even there is no connection between C_2 and C_3 in the emerging e-commerce site.

We further integrate the similarity information captured by AmpSim with a CMF model such that the latent factors of similar entities would be refined w.r.t. the geometric structure. As a consequence, non-overlapping entities can also benefit from the transferred knowledge through the latent factor refinement. Hence, the transferred knowledge would not be biased to the more active ones but can be iteratively propagated across networks to fit the emerging business. Through extensive experiments on real-world datasets, we demonstrate that our proposed model significantly outperforms other state-of-the-art collaborative filtering algorithms in addressing item recommendation for emerging businesses.

2 Preliminaries

In this section, we present the preliminaries and the problem formulation of this study. Table 1 lists the main notations we use through this paper.

We extend the definition of Heterogenous Information Network [Sun *et al.*, 2011] to take the link attributes augmented with the links into account.

Definition 1. Augmented Heterogeneous Information Network (AHIN): An AHIN can be represented as $G = (\mathcal{V}, \mathcal{E}, \mathcal{Y})$, where each entity $v \in \mathcal{V}$ belongs to an entity type $\tau(v) \in T$, and each link $l \in \mathcal{E}$ belongs to a relation type $\psi(l) \in R$ and may have an augmented attribute y, which belongs to an link attribute type $\omega(y) \in \mathcal{W}$.

An online business can be modeled using an AHIN: $G = (\mathcal{V}, \mathcal{E}, \mathcal{Y})$, where the user set $\mathcal{U}$ and item set $\mathcal{I}$ are subsets of entities (i.e., $\mathcal{U}, \mathcal{I} \subset \mathcal{V}$), the existing feedbacks of items given by users are subsets of links (i.e., $\mathcal{U} \times \mathcal{I} \subset \mathcal{E}$). The attributes augmented with the feedbacks (such as the user-item ratings $\mathbf{Y}$, where each entry $\mathbf{Y}_{ui}$ corresponds to the rating of user $u \in \mathcal{U}$ on item $i \in \mathcal{I}$) are kept as link attributes in G

Table 1: Notation summary

Symbol	Description
u, i	user, item
$\mathbf{Y}, \mathbf{I}$	user-item rating and indicator matrices
$\mathbf{P}, \mathbf{Q}$	low rank representations of users and items
G, S_G	augmented information network and its schema
$\mathcal{V}, \mathcal{E}, \mathcal{W}$	set of entities, links and link attributes
$\mathcal{A}^{(s,t)}$	set of anchor links across network $G^{(s)}$ and $G^{(t)}$
T, R	entity type and link type
$\overline{G}$	aligned networks
$\mathcal{P}$	augmented meta path

(i.e., $\mathbf{Y} \subset \mathcal{Y}$). An emerging business is a business in which the average number of ratings is lower than a threshold, i.e., $AvgDeg(\mathbf{Y}) = \frac{|\mathbf{Y}|}{|\mathcal{U}|} < \epsilon$.

If pairs of different networks share some common entities, then these networks are called aligned networks.

Definition 2. Aligned Networks: Aligned networks can be formulated as $\overline{G} = ((G^{(1)}, G^{(2)}, \ldots, G^{(\Pi)}), \mathcal{A})$, where $\mathcal{A} = \bigcup_{s,t} \mathcal{A}^{(s,t)}, 1 \le s < t \le \Pi$. $\mathcal{A}^{(s,t)}$ is the set of undirected anchor links $(v_p^{(s)}, v_q^{(t)})$ between entities across network $G^{(s)}$ and $G^{(t)}$, where $v_p^{(s)} \in \mathcal{V}^{(s)}$ and $v_q^{(t)} \in \mathcal{V}^{(t)}$.

We can depict the network schema of an AHIN G as $S_G = (T, R, W)$, and the network schema of of an aligned networks $\overline{G}$ as $S_{\overline{G}} = (\overline{T}, \overline{R}, \overline{W})$, where $\overline{T} = \bigcup_{\pi} T^{(\pi)}$ is the union of entity types and $\overline{R} = \bigcup_{\pi} R^{(\pi)} \bigcup R^A$ is the union of link types (including the anchor link relation R^A), and $\overline{W} = \bigcup_{\pi} W^{(\pi)}$ is the union of link attribute types, respectively. The network schema of the aligned networks in Figure 1, for instance, is shown in Figure 2, where rectangles, diamonds and circles represent the entity types, relation types and link attribute types, respectively.

For the sake of simplicity, we focus on one target emerging business network $G^{(t)}$ and one source network $G^{(s)}$ (i.e., $\Pi = \{s, t\}$), which can be easily extended to multiple related business. The task in this paper is to predict the missing rating in the target business $G^{(t)}$. The predicted rating of user u on item j is denoted by $\hat{\mathbf{Y}}_{uj}^{(t)}$. Given aligned networks $\overline{G}$, we aims at providing a predictive function $f : \overline{G} \to \hat{\mathbf{Y}}^{(t)}$, such that the prediction errors are minimized.

3 Augmented Meta Path-based Similarity

To measure entity similarity by taking both the linkage structure and link attributes into account, we define the augmented meta paths as follows:

Definition 3. Augmented meta path (AMP): Given the network schema $S_G = (T, R, W)$, an AMP $\mathcal{P}$ is a sequence of relations augmented with link attributes between entities, and is denoted in the form of $T_1 \xrightarrow{R_1[W_1]} T_2 \xrightarrow{R_2[W_2]} \ldots \xrightarrow{R_{l-1}[W_{l-1}]} T_l$, where R_i is the relation type between entity type T_i and T_{i+1}, and W_i is the augmented link attribute type of R_i if exists.
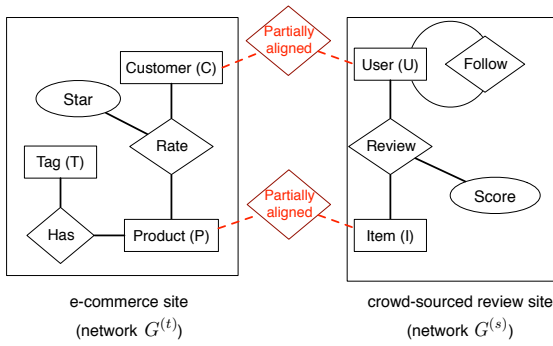
Figure 2: Schema of an e-commerce site partially aligned with a crowd-sourced review site.

For simplicity, we will use the first letter of each entity to represent that entity, the opposite direction of the arrow to represent the reverse relation, ignore the relation type if there is no ambiguity, and ignore the link attribute type if it is simply the count of connections. If an AMP $\mathcal{P}$ does not involve any anchor link relations R^A, we call it as an intra-network AMP. Otherwise, we call it as an inter-network AMP.

Example 1. *Intra-network AMP:* The products rated by the same customer can be captured by $P \xleftarrow{\text{[star]}} C \xrightarrow{\text{[star]}} P$.

Example 2. *Inter-network AMP:* The products in $G^{(t)}$ that are reviewed by the same users in $G^{(s)}$ can be observed by $P \leftrightarrow I \xleftarrow{\text{[score]}} U \xrightarrow{\text{[score]}} I \leftrightarrow P$, where $\leftrightarrow$ denotes the anchor links across $G^{(s)}$ and $G^{(t)}$.

Inspired by the Homophily Principle [McPherson *et al.*, 2001] – two entities are considered to be similar if they are related to similar neighbors – we propose to measure the similarity of two entities by their neighbors' similarities. However, an AMP may consist of multiple types of relations, and different link attributes are hard to be compared. To provide a general similarity measure, we apply a simple trick to normalize the link attributes in the AMP.

Let first assume that in the AMP the link attributes appear in pairs (i.e., the number of the same type of link attributes is even). We can replace the link attribute $\mathbf{Y}$ by the normalized value $\mathbf{M}$ using following equation:

$$\mathbf{M}_{ui} = \frac{Y_{ui} - b_i}{\sqrt{\sum_j (Y_{uj} - b_j)^2}}, \qquad (1)$$

where b_i is a bias term for the entity i. For the link attributes that are bounded within a certain range (e.g., a rating can be bounded between 1 and 5), we set $b_i = \overline{Y_i}$ the average value of entity i. For the other type of attributes, we set $b_i = 0$. It is not hard to see that the multiplication of a pair of normalized link attributes (i.e., $\mathbf{MM}^T$) equals to adjusted cosine similarity if $b_i \neq 0$ and equals to cosine similarity if $b_i = 0$. For the link attributes that do not appear in pairs, there is no way to compare the link attributes with others. Thus, we replace these attributes by the count of the connections and normalize them using Equation 1 with $b_i = 0$.

Considering the normalized link attributes and the direction of AMPs, we formulate the similarity measure as:

Definition 4. AmpSim: An augmented meta path-based similarity measure between v_a and v_b based on path $\mathcal{P}$ is:

$$s(v_a, v_b | \mathcal{P}) = \frac{[\prod_{i=1}^l \mathbf{M}_i]_{ab} + [\prod_{i=l}^1 \mathbf{M}_i^T]_{ba}}{[\prod_{i=1}^l \mathbf{M}_i]_{a*} + [\prod_{i=l}^1 \mathbf{M}_i^T]_{b*}} \in [0, 1],$$

where $\prod_{i=1}^l \mathbf{M}_i$ denotes the product of the normalized link attributes upon $\mathcal{P}$, $[M]_{ab}$ means the entry (a, b) in $\mathbf{M}$, and $[M]_{a*}$ means the ath row in $\mathbf{M}$. Since *AmpSim* can be computed in matrix form, the time complexity of computing AmpSim equals to that of matrix multiplications. Note that when $\mathcal{P}$ is symmetric and all the link attributes are the count of connections, *AmpSim* is equivalent to *PathSim* [Sun *et al.*, 2011]. Hence, AmpSim can be seen as a generalized version of *PathSim* on AMPs.

Different AMPs capture the similarity between entities in different aspects and overall similarity between entities can be obtained by aggregating information from all these AMPs. Without loss of generality, we choose *tanh* as the aggregation function, the overall similarity between entity v_a and v_b can be represented as

$$S(v_a, v_b) = \frac{tanh(\sum_i w_i S_i(v_a, v_b))}{tanh(1)} \in [0, 1], \qquad (2)$$

where S_i denotes the *AmpSim* upon the AMP $\mathcal{P}_i$, the value of w_i denotes the weight of $\mathcal{P}_i$, and $\sum_i w_i = 1$.

4 Recommendation Model

In transfer learning based collaborative filtering, collective matrix factorization (CMF) [Singh and Gordon, 2008] is proposed to estimate the low-rank representations as follows:

$$\min_{\mathbf{P}^{(\pi)}, \mathbf{Q}^{(\pi)}, \pi \in \{s,t\}} \sum_{\pi \in \{s,t\}} ||\mathbf{I}^{(\pi)} \odot (\mathbf{Y}^{(\pi)} - \mathbf{P}^{(\pi)}\mathbf{Q}^{(\pi)^T})||_F^2,$$

where $\odot$ is the Hadamard (element-wise) product, $|| \cdot ||_F$ stands for Frobenius norm, and $\mathbf{I}^{(\pi)}$ is an indicator matrix. $\mathbf{I}_{ui}^{(\pi)} = 1$ if $\mathbf{Y}_{ui}^{(\pi)}$ is observed, and otherwise $\mathbf{I}_{ui}^{(\pi)} = 0$. $\mathbf{P}^{(\pi)} = [\mathbf{p}_1^{(\pi)}; \mathbf{p}_2^{(\pi)}; \ldots; \mathbf{p}_{n_\pi}^{(\pi)}] \in \mathbb{R}^{n_\pi \times k}$ and $\mathbf{Q}^{(\pi)} = [\mathbf{q}_1^{(\pi)}; \mathbf{q}_2^{(\pi)}; \ldots; \mathbf{q}_{m_\pi}^{(\pi)}] \in \mathbb{R}^{m_\pi \times k}$ are low-rank representation of users and items. n_π and m_π are the number of users and items in network $G^{(\pi)}$, and k is the parameter that estimate the rank. The key idea of CMF is to share parameters among factors when an entity participates in multiple relations. In [Singh and Gordon, 2008], for instance, the factor matrices of items are assumed to be the same (i.e., $\mathbf{Q}^{(s)} = \mathbf{Q}^{(t)}$).

We formulate our model as a constrained collective matrix factorization, where three soft constraints are involved:

- **Non-negativity**: The factor matrices $\{\mathbf{P}^{(\pi)}\}, \{\mathbf{Q}^{(\pi)}\}$ contain only nonnegative entries, since we only focus on positive interactions between users and items.

- **Geometric closeness**: The latent factors of similar entities should be close w.r.t. the geometric structure. Preserving the geometric structure in the target domain can be achieved by the geometric regularization [Cai *et al.*, 2011]:

$$\mathcal{R}_G(\mathbf{P}) = \frac{1}{2} \sum_{u,v} \mathbf{S}_{uv} ||\mathbf{p}_u - \mathbf{p}_v||_2^2,$$

where $\mathbf{S}$ is computed by Equation 2.

- **Alignment constraint**: The latent factors of aligned entities should be close. Let (i, g_i) be the anchor link between entity i in the target network $G^{(t)}$ and entity g_i in the source network $G^{(s)}$. The difference of their latent factors should be minimized:

$$\mathcal{R}_A(\mathbf{Q}^{(t)}, \mathbf{Q}^{(s)}) = \frac{1}{2} \sum_{(i,g_i) \in \mathcal{A}^{(s,t)}} ||\mathbf{q}_i^{(t)} - \mathbf{q}_{g_i}^{(s)}||_2^2.$$

Integrating the alignment regularization $\mathcal{R}_A$ and the geometric regularization $\mathcal{R}_G$ seamlessly would enjoy the intrinsic mutual reinforcement learning: 1) with $\mathcal{R}_A$, knowledge can be transferred between businesses through the common latent factors of overlapping entities; 2) with $\mathcal{R}_G$, the latent factors can be refined for similar entities to fit the geometric closeness within the emerging business. The objective function of our model can be formulated as follows:

$$\min_{\substack{\mathbf{P}^{(\pi)}, \mathbf{Q}^{(\pi)} \geq 0, \\ \pi \in \{s,t\}}} \mathcal{J} = \sum_{\pi \in \{s,t\}} ||\mathbf{I}^{(\pi)} \odot (\mathbf{Y}^{(\pi)} - \mathbf{P}^{(\pi)} \mathbf{Q}^{(\pi)^T})||_F^2$$
$$+ \frac{\alpha}{2} \sum_{\pi \in \{s,t\}} (||\mathbf{P}^{(\pi)}||_F^2 + ||\mathbf{Q}^{(\pi)}||_F^2)$$
$$+ \beta (\mathcal{R}_A(\mathbf{P}^{(t)}, \mathbf{P}^{(s)}) + \mathcal{R}_A(\mathbf{Q}^{(t)}, \mathbf{Q}^{(s)}))$$
$$+ \lambda (\mathcal{R}_G(\mathbf{P}^{(t)}) + \mathcal{R}_G(\mathbf{Q}^{(t)})), \qquad (3)$$

where α controls the regularization to avoid over-fitting when learning $\{\mathbf{P}^{(\pi)}\}, \{\mathbf{Q}^{(\pi)}\}$, β controls the trade-off between source network and target network and λ controls the importance of geometric closeness. Since we focus on improving recommendation in the target business, the geometric regularization $\mathcal{R}_G$ is only applied to $\mathbf{P}^{(t)}$ and $\mathbf{Q}^{(t)}$.

To ease the subsequent derivation, we rewrite the geometric terms into trace form. From the similarity matrix $\mathbf{S}$ on users, we define the diagonal matrix $\mathbf{D}$ whose elements $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{ij}$, and the Laplacian matrix $\mathbf{L}_P = \mathbf{D} - \mathbf{S}$. Then $\mathcal{R}_G(\mathbf{P})$ can be reduced into the trace form:

$$\sum_{u,v} \mathbf{S}_{uv} ||\mathbf{p}_u - \mathbf{p}_v||^2 = Tr(\mathbf{P}^T(\mathbf{D} - \mathbf{S})\mathbf{P}) = Tr(\mathbf{P}^T \mathbf{L}_P \mathbf{P}).$$

Similarly, given the similarity matrix $\mathbf{S}'$ on items, $\mathcal{R}_G(Q)$ can also be reduced into the trace form $Tr(\mathbf{Q}^T \mathbf{L}_Q \mathbf{Q})$, where $\mathbf{L}_Q = \mathbf{D}' - \mathbf{S}'$ and $\mathbf{D}'_{ii} = \sum_j \mathbf{S}'_{ij}$.

For solving Equation 3, we derive the multiplicative updating rules w.r.t. $\{\mathbf{P}^{(\pi)}\}, \{\mathbf{Q}^{(\pi)}\}$ in the rest of this section.

Let $\mathbf{A}$ be the user mapping matrix that map users from network $\mathcal{G}^{(t)}$ to network $\mathcal{G}^{(s)}$, i.e., $\mathbf{A}_{u,g_u} = 1$ if $(u, g_u) \in \mathcal{A}^{(t,s)}$, zero otherwise. The partial derivatives of $\mathcal{J}$ in Equation 3 w.r.t. $\{\mathbf{P}^{(\pi)}\}$ are:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{P}^{(s)}} = -(\mathbf{I}^{(s)} \odot \mathbf{Y}^{(s)})\mathbf{Q}^{(s)} + \mathbf{I}^{(s)} \odot (\mathbf{P}^{(s)}\mathbf{Q}^{(s)^T})\mathbf{Q}^{(s)}$$
$$+ \alpha \mathbf{P}^{(s)} + \beta(\mathbf{A}^T \mathbf{A} \mathbf{P}^{(s)} - \mathbf{A}^T \mathbf{P}^{(t)})$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{P}^{(t)}} = -(\mathbf{I}^{(t)} \odot \mathbf{Y}^{(t)})\mathbf{Q}^{(t)} + \mathbf{I}^{(t)} \odot (\mathbf{P}^{(t)}\mathbf{Q}^{(t)^T})\mathbf{Q}^{(t)}$$
$$+ \alpha \mathbf{P}^{(t)} + \beta(\mathbf{A}\mathbf{A}^T \mathbf{P}^{(t)} - \mathbf{A}\mathbf{P}^{(s)}) + \lambda \mathbf{L}_P \mathbf{P}^{(t)}$$

Using the Karush-Kuhn-Tucker (KKT) complementarity conditions, we can obtain the following updating rules:

$$\mathbf{P}^{(s)} = \mathbf{P}^{(s)} \odot$$
$$\sqrt{\frac{(\mathbf{I}^{(s)} \odot \mathbf{Y}^{(s)})\mathbf{Q}^{(s)} + \beta \mathbf{A}^T \mathbf{P}^{(t)}}{\mathbf{I}^{(s)} \odot (\mathbf{P}^{(s)}\mathbf{Q}^{(s)^T})\mathbf{Q}^{(s)} + \alpha \mathbf{P}^{(s)} + \beta \mathbf{A}^T \mathbf{A} \mathbf{P}^{(s)}}},$$

$$\mathbf{P}^{(t)} = \mathbf{P}^{(t)} \odot$$
$$\sqrt{\frac{(\mathbf{I}^{(t)} \odot \mathbf{Y}^{(t)})\mathbf{Q}^{(t)} + \beta \mathbf{A}\mathbf{P}^{(s)} + \lambda \mathbf{L}_P^- \mathbf{P}^{(t)}}{\mathbf{I}^{(t)} \odot (\mathbf{P}^{(t)}\mathbf{Q}^{(t)^T})\mathbf{Q}^{(t)} + \alpha \mathbf{P}^{(t)} + \beta \mathbf{A}\mathbf{A}^T \mathbf{P}^{(t)} + \lambda \mathbf{L}_P^+ \mathbf{P}^{(t)}}},$$

where $\mathbf{L}_P = \mathbf{L}_P^+ - \mathbf{L}_P^-$, $[\mathbf{L}_P^+]_{ij} = (|[\mathbf{L}_P]_{ij}| + [\mathbf{L}_P]_{ij})/2 \geq 0$ and $[\mathbf{L}_P^-]_{ij} = (|[\mathbf{L}_P]_{ij}| - [\mathbf{L}_P]_{ij})/2 \geq 0$.

Similarly, let $\mathbf{B}$ be the item mapping matrix whose element $\mathbf{B}_{i,g_i} = 1$ if $(i, g_i) \in \mathcal{A}^{(t,s)}$, zero otherwise. We can obtain the following updating rules from the derivatives of $\mathcal{J}$ w.r.t. $\{\mathbf{Q}^{(\pi)}\}$:

$$\mathbf{Q}^{(s)} = \mathbf{Q}^{(s)} \odot$$
$$\sqrt{\frac{(\mathbf{I}^{(s)} \odot \mathbf{Y}^{(s)})^T \mathbf{P}^{(s)} + \beta \mathbf{B}^T \mathbf{Q}^{(t)}}{(\mathbf{I}^{(s)} \odot (\mathbf{P}^{(s)}\mathbf{Q}^{(s)^T}))^T \mathbf{P}^{(s)} + \alpha \mathbf{Q}^{(s)} + \beta \mathbf{B}^T \mathbf{B} \mathbf{Q}^{(s)}}},$$

$$\mathbf{Q}^{(t)} = \mathbf{Q}^{(t)} \odot$$
$$\sqrt{\frac{(\mathbf{I}^{(t)} \odot \mathbf{Y}^{(t)})^T \mathbf{P}^{(t)} + \beta \mathbf{B}^T \mathbf{Q}^{(s)} + \lambda \mathbf{L}_Q^- \mathbf{Q}^{(t)}}{(\mathbf{I}^{(t)} \odot (\mathbf{P}^{(t)}\mathbf{Q}^{(t)^T}))^T \mathbf{P}^{(t)} + \alpha \mathbf{Q}^{(t)} + \beta \mathbf{B}^T \mathbf{B} \mathbf{Q}^{(t)} + \lambda \mathbf{L}_Q^+ \mathbf{Q}^{(t)}}},$$

where $\mathbf{L}_Q = \mathbf{L}_Q^+ - \mathbf{L}_Q^-$, $[\mathbf{L}_Q^+]_{ij} = (|[\mathbf{L}_Q]_{ij}| + [\mathbf{L}_Q]_{ij})/2 \geq 0$ and $[\mathbf{L}_Q^-]_{ij} = (|[\mathbf{L}_Q]_{ij}| - [\mathbf{L}_Q]_{ij})/2 \geq 0$.

Given the learned latent factors, the missing ratings in $\mathbf{Y}^{(t)}$ are predicted as $\hat{\mathbf{Y}}^{(t)} = \mathbf{P}^{(t)} \mathbf{Q}^{(t)^T}$.

5 Experiments
5.1 Experimental Setup

We evaluate our proposed recommendation model on two real-world datasets: Yelp[1] and Epinions [Tang *et al.*, 2012]. For both datasets, we filter out all the locations/items with less than 5 ratings and all the users who only have 1 rating. The statistics of the datasets after filtering are listed in Table 2.

The Epinions dataset is used for testing the scenario of an emerging online business that shares some users with a developed online business. As in [Li and Lin, 2014], we partition the dataset into two parts with partially overlapping users and items. One part consists of ratings given by 80% users, which serves as the source business. The other part consists of ratings given by 20% users, which serves as the target business. In this task, 20% users and all the items in the target business are overlapping with the source business. The Yelp dataset is used for testing the scenario of a mature online business that starts to branch into new geographic areas. There are $176,736$ ratings on $6,317$ locations given by $19,464$ users in Arizona, and $53,682$ ratings on $2,150$ locations given by $11,476$ users in Nevada. We consider Arizona as the source business and Nevada as the target business. $4,322$ users are shared in both businesses but the locations are disjoint.

For each target business, we conduct the experiments using 5-fold cross-validation: one fold is used as training data, the

---

[1] http://www.yelp.com/dataset_challenge/

Table 2: Statistics of Datasets

Name	#users	#items/locations	#tags	#ratings	#social links
Epinions	$21,740$	$31,678$	27	$541,108$	$344,286$
Yelp	$26,618$	$8,467$	770	$230,418$	$183,765$

Table 3: Augmented meta paths used in the experiment.

Intra-network augmented meta path
$(C \xrightarrow{[star]} P \xleftarrow{[star]} C),(P \xleftarrow{[star]} C \xrightarrow{[star]} P), (P \rightarrow T \leftarrow P)$
$(C \xrightarrow{[star]} P \xrightarrow{[star]} C \xrightarrow{[star]} P \xleftarrow{[star]} C)$
$(P \xleftarrow{[star]} C \xrightarrow{[star]} P \xrightarrow{[star]} C \xrightarrow{[star]} P)$
Inter-network augmented meta path
$(P \leftrightarrow I \xleftarrow{[score]} U \xrightarrow{[score]} I \leftrightarrow P)$
$(C \leftrightarrow U \xrightarrow{[score]} I \xleftarrow{[score]} U \leftrightarrow C), (C \leftrightarrow U \leftrightarrow U \leftrightarrow C)$
C, P and T denote "customer", "product" and "tag" in $G^{(t)}$, respectively.
U and I denote "user" and "item" in $G^{(s)}$, respectively.
$\leftrightarrow$ denotes the anchor links across $G^{(s)}$ and $G^{(t)}$.

remaining folds are used as testing data. We report the average results of 5-fold cross validation on the datasets. Since we are interested in the cold start problem, we select the cold start users, who have less than 5 ratings in the training set, in each fold and report their results separately. We use Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) as the evaluation metric to measure the prediction quality. Both metrics have been widely used in the recommendation problem [Zhang *et al.*, 2006; Yu *et al.*, 2013; Luo *et al.*, 2014] and the smaller is the value of each criterion, the better is the performance.

5.2 Comparison methods

We compare our proposed model with several state-of-the-art recommendation models. First, we implement the models that utilize heterogeneous relationships to capture similarity information as geometric regularization for improving recommendation performance. Weighted nonnegative matrix factorization (**WNMF** [Zhang *et al.*, 2006]) on the target rating matrix is utilized as the baseline method. In the following methods, different similarity measures are introduced for constructing the geometric regularization and incorporating into the WNMF model.

- **Hete-MF** [Yu *et al.*, 2013]: uses *PathSim* to measure the similarity between items.

- **Hete-CF** [Luo *et al.*, 2014]: extends Hete-MF by also considering the relationship between users.

- **Hete-PRW**: We implement *pairwise random walk* to measure the similarity between users/items.

- **Amp-MF**: We use our proposed similarity measure *AmpSim* to compute the similarity between users/items.

In the experiments, we construct aligned networks for both datasets based on the network schema in Figure 2 and utilize eight different similarity semantics listed in Table 3 to capture the similarity information for the above models. For simplicity, the weights of different AMPs in Equation 2 are assigned with identical values, i.e., $\omega = [\frac{1}{8}, , \frac{1}{8}]$.

Next, we compare with the models that learn common latent factors from multiple relative matrices.

- **CMF** [Singh and Gordon, 2008]: Collective matrix factorization with alignment regularization is used as the state-of-the-art transfer learning model with cross-domain entity correspondences.

- **RMGM** [Li *et al.*, 2009b]: Rating-matrix generative model is used as the state-of-the-art model without entity correspondences.

- **Amp-CMF**: Our proposed recommendation model.

Parameters of baselines are set to be consistent with the values recommended in their corresponding papers. For each method, we randomly initialize the latent factors and set the maximum number of iterations as 100. For **RMGM**, the latent dimensionality k is set as the number of clusters, the default choice for most kernel-based approaches. The sparsity tradeoff parameter α is fixed as 0.1 for both datasets. We set the similarity tradeoff parameter $\lambda = 1$ as in [Yu *et al.*, 2013; Luo *et al.*, 2014] and tune the alignment tradeoff parameter β by searching the grid of $\{10^{-5}, \cdots, 10^3\}$.

5.3 Performance Analysis

Table 4 and Table 5 present the RMSE and MAE results of all methods, with different dimensionality settings ($k = 10$ and 20), on both datasets. The column Overall and Cold Start report the performance on the whole testing user set and the cold start user set, respectively. The best results in each experimental setting are listed in bold.

We first investigate the performance of the models with geometric regularization. It can be seen that Hete-CF performs better than WNMF and Hete-MF, likely due to the fact that incorporating more similarity information can alleviate the data sparsity issue and improve the prediction accuracy. For all cases, Amp-MF beats all baseline methods. We believe this is because the more accurate similarity measure can usually lead to the better recommendation performance. Our proposed similarity measure, AmpSim, is capable of leveraging the relationship attributes to find the most similar entities.

Next, we compare the models that learn common latent factors from multiple relative matrices. CMF preforms better than RMGM in the Epinions dataset while the results are opposite in the Yelp dataset. The major reason is that CMF requires entity correspondences for knowledge transfer. In the Epinions dataset, the items in the target business are contained in the source business, but in the Yelp dataset the items (locations) are disjoint. The lack of overlapping items makes CMF not able to transfer the common information of items from the source business to the target business. Furthermore, we can see that Amp-CMF consistently outperforms all the other comparison models. This confirms our assumption that integrating the similarity information with the CMF model would enjoy the intrinsic mutual reinforcement and boost the recommendation quality.

In general, most of the models with higher dimensionality ($k = 20$) perform slightly better than that with lower dimensionality ($k = 10$). Besides, all models have higher prediction errors on the cold start user sets than on the whole testing user sets. Noteworthily, Amp-CMF also outperforms all the other baseline methods for the cold start users. This confirms

Table 4: Performance comparison (mean±std) on the Epinions dataset

Method	Overall				Cold Start			
	RMSE		MAE		RMSE		MAE	
	$k = 10$	$k = 20$	$k = 10$	$k = 20$	$k = 10$	$k = 20$	$k = 10$	$k = 20$
WNMF	1.551 ± 0.009	1.533 ± 0.009	1.156 ± 0.008	1.153 ± 0.009	1.596 ± 0.016	1.594 ± 0.016	1.219 ± 0.016	1.218 ± 0.013
Hete-MF	1.402 ± 0.010	1.398 ± 0.009	1.034 ± 0.009	1.030 ± 0.008	1.462 ± 0.017	1.454 ± 0.014	1.106 ± 0.016	1.101 ± 0.013
Hete-CF	1.148 ± 0.008	1.141 ± 0.008	0.908 ± 0.007	0.901 ± 0.007	1.211 ± 0.011	1.201 ± 0.011	0.961 ± 0.009	0.955 ± 0.009
Hete-PRW	1.395 ± 0.009	1.392 ± 0.009	1.039 ± 0.005	1.030 ± 0.005	1.434 ± 0.009	1.428 ± 0.009	1.072 ± 0.005	1.066 ± 0.005
Amp-MF	$\mathbf{1.099 \pm 0.009}$	$\mathbf{1.097 \pm 0.009}$	$\mathbf{0.869 \pm 0.005}$	$\mathbf{0.868 \pm 0.005}$	$\mathbf{1.131 \pm 0.009}$	$\mathbf{1.128 \pm 0.009}$	$\mathbf{0.899 \pm 0.005}$	$\mathbf{0.897 \pm 0.005}$
CMF	1.152 ± 0.007	1.143 ± 0.007	0.870 ± 0.004	0.868 ± 0.005	1.198 ± 0.012	1.185 ± 0.010	0.902 ± 0.009	0.899 ± 0.009
RMGM	1.246 ± 0.008	1.242 ± 0.010	0.989 ± 0.005	0.983 ± 0.007	1.271 ± 0.005	1.266 ± 0.009	1.013 ± 0.002	1.014 ± 0.006
Amp-CMF	$\mathbf{1.097 \pm 0.009}$	$\mathbf{1.095 \pm 0.009}$	$\mathbf{0.867 \pm 0.005}$	$\mathbf{0.866 \pm 0.005}$	$\mathbf{1.129 \pm 0.009}$	$\mathbf{1.127 \pm 0.009}$	$\mathbf{0.898 \pm 0.005}$	$\mathbf{0.896 \pm 0.005}$

Table 5: Performance comparison (mean±std) on the Yelp dataset

Method	Overall				Cold Start			
	RMSE		MAE		RMSE		MAE	
	$k = 10$	$k = 20$	$k = 10$	$k = 20$	$k = 10$	$k = 20$	$k = 10$	$k = 20$
WNMF	1.446 ± 0.009	1.429 ± 0.009	1.097 ± 0.006	1.083 ± 0.005	1.535 ± 0.014	1.520 ± 0.013	1.184 ± 0.005	1.170 ± 0.005
Hete-MF	1.429 ± 0.009	1.351 ± 0.009	1.086 ± 0.005	1.006 ± 0.005	1.518 ± 0.012	1.492 ± 0.011	1.171 ± 0.005	1.148 ± 0.005
Hete-CF	1.305 ± 0.008	1.199 ± 0.008	0.957 ± 0.005	0.907 ± 0.005	1.378 ± 0.011	1.228 ± 0.009	1.017 ± 0.005	0.935 ± 0.005
Hete-PRW	1.343 ± 0.008	1.313 ± 0.008	1.018 ± 0.005	0.991 ± 0.005	1.414 ± 0.008	1.382 ± 0.008	1.088 ± 0.005	1.059 ± 0.005
Amp-MF	$\mathbf{1.191 \pm 0.009}$	$\mathbf{1.187 \pm 0.009}$	$\mathbf{0.899 \pm 0.005}$	$\mathbf{0.897 \pm 0.005}$	$\mathbf{1.219 \pm 0.008}$	$\mathbf{1.215 \pm 0.008}$	$\mathbf{0.928 \pm 0.005}$	$\mathbf{0.925 \pm 0.005}$
CMF	1.294 ± 0.009	1.274 ± 0.010	0.966 ± 0.005	0.949 ± 0.006	1.349 ± 0.012	1.329 ± 0.012	1.015 ± 0.005	0.998 ± 0.006
RMGM	1.240 ± 0.009	1.238 ± 0.009	0.925 ± 0.005	0.902 ± 0.004	1.316 ± 0.009	1.295 ± 0.009	0.995 ± 0.004	0.974 ± 0.004
Amp-CMF	$\mathbf{1.134 \pm 0.009}$	$\mathbf{1.127 \pm 0.009}$	$\mathbf{0.854 \pm 0.005}$	$\mathbf{0.847 \pm 0.005}$	$\mathbf{1.148 \pm 0.009}$	$\mathbf{1.139 \pm 0.009}$	$\mathbf{0.875 \pm 0.006}$	$\mathbf{0.865 \pm 0.006}$

the effectiveness of Amp-CMF in addressing item recommendation for emerging businesses.

6 Related Work

Transfer learning based on CMF has been proposed to utilized information shared by related domains to learn latent factors for better recommendations. The first category of such methods assume that there are certain overlapping users and/or items across multiple domains and align the latent factors of the overlapping users and/or items [Singh and Gordon, 2008; Long *et al.*, 2010]. Another approach is to enforce the similarity of cluster-level preference patterns in two related domains, without assuming cross-domain entity correspondences. Codebook-based-transfer (CBF) [Li *et al.*, 2009a] proposed to align the two kernel matrices of users from two domains using a co-clustering process. Rating-matrix generative model (RMGM) relaxes the constraints in CBT from hard clustering to soft clustering by using a probabilistic graphical model. Although the former approach is more effective for knowledge transfer in general, it can be biased to the overlapping portions. One possible way to handle the above issue is to preserve geometric closeness between similar entities [Wang and Mahadevan, 2011; Long *et al.*, 2014]. However, it may not be trivial to find reliable similarity information in the emerging domain.

When building a recommendation system, it is crucial to choose a good way to measure the entity similarity. A number of studies leverage linkage structure of a network for measuring similarity, e.g., *personalized PageRank* [Jeh and Widom, 2003] and *SimRank* [Jeh and Widom, 2002], while they are defined on a homogeneous network or a bipartite network. However, in real-world applications, there are lots of heterogeneous networks. Meta path was introduced in [Sun *et al.*, 2011; Shi *et al.*, 2012] to measure the similarity in heterogeneous networks. A series of meta path-based similarity measures are proposed for either the same type of entities [Sun *et al.*, 2011] or different type of entities [Shi *et al.*, 2014; Cao *et al.*, 2014; Shi *et al.*, 2015]. Recently, researchers have been aware of the importance of heterogeneous information for recommendation. Yu et al. [Yu *et al.*, 2013] proposed an implicit feedback recommendation model with the similarity information extracted from heterogeneous network. Luo et al. [Luo *et al.*, 2014] proposed a collaborative filtering-based social recommendation method, called Hete-CF, using heterogeneous relations. Vahedian [Vahedian, 2014] proposed the WHyLDR approach for multiple recommendation tasks, which combines heterogeneous information with a linear-weighted hybrid model. These methods usually treat the relationships between entities as binary connections. However, plentiful attributes are usually attached to the relationships, e.g., the rating of an item given by a user and the timestamp of a post. As demonstrated in the experiments, discarding these important attributes can lead to a degraded performance.

7 Conclusion

In this paper, we have proposed a novel similarity measure called *AmpSim* that can judiciously capture the rich similarity semantics between entities by taking both the linkage structures and the augmented link attributes into account. We further incorporated the similarity information captured by AmpSim in a constrained collective matrix factorization model. Extensively experiments on real-world datasets have demonstrated that our proposed model significantly outperforms other state-of-the-art collaborative filtering algorithms in addressing item recommendation for emerging businesses.

Acknowledgments

References

[Cai *et al.*, 2011] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1548–1560, 2011.

[Cao *et al.*, 2014] Bokai Cao, Xiangnan Kong, and Philip S. Yu. Collective prediction of multiple types of links in heterogeneous information networks. In *ICDM*, pages 50–59, 2014.

[Jeh and Widom, 2002] Glen Jeh and Jennifer Widom. Simrank: a measure of structural-context similarity. In *SIGKDD*, pages 538–543, 2002.

[Jeh and Widom, 2003] Glen Jeh and Jennifer Widom. Scaling personalized web search. In *WWW*, pages 271–279, 2003.

[Kong *et al.*, 2013] Xiangnan Kong, Jiawei Zhang, and Philip S. Yu. Inferring anchor links across multiple heterogeneous social networks. In *CIKM*, pages 179–188, 2013.

[Koren, 2008] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *SIGKDD*, pages 426–434, 2008.

[Li and Lin, 2014] Chung-Yi Li and Shou-De Lin. Matching users and items across domains to improve the recommendation quality. In *SIGKDD*, pages 801–810, 2014.

[Li *et al.*, 2009a] Bin Li, Qiang Yang, and Xiangyang Xue. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, pages 2052–2057, 2009.

[Li *et al.*, 2009b] Bin Li, Qiang Yang, and Xiangyang Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *ICML*, pages 617–624, 2009.

[Long *et al.*, 2010] Mingsheng Long, Wei Cheng, Xiaoming Jin, Jianmin Wang, and Dou Shen. Transfer learning via cluster correspondence inference. In *ICDM*, pages 917–922, 2010.

[Long *et al.*, 2014] Mingsheng Long, Jianmin Wang, Guiguang Ding, Dou Shen, and Qiang Yang. Transfer learning with graph co-regularization. *TKDE*, 26(7):1805–1818, 2014.

[Lu *et al.*, 2014] Chun-Ta Lu, Hong-Han Shuai, and Philip S. Yu. Identifying your customers in social networks. In *CIKM*, pages 391–400, 2014.

[Luo *et al.*, 2014] Chen Luo, Wei Pang, Zhe Wang, and Chenghua Lin. Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations. In *ICDM*, pages 917–922, 2014.

[McPherson *et al.*, 2001] Miller McPherson, Lynn S. Lovin, and James M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[Rong *et al.*, 2014] Yu Rong, Xiao Wen, and Hong Cheng. A monte carlo algorithm for cold start recommendation. In *WWW*, pages 327–336, 2014.

[Shi *et al.*, 2012] Chuan Shi, Chong Zhou, Xiangnan Kong, Philip S. Yu, Gang Liu, and Bai Wang. Heterecom: a semantic-based recommendation systemin heterogeneous networks. In *SIGKDD*, pages 1552–1555, 2012.

[Shi *et al.*, 2014] Chuan Shi, Xiangnan Kong, Yue Huang, Philip S. Yu, and Bin Wu. Hetesim: A general framework for relevance measure in heterogeneous networks. *TKDE*, 26(10):2479–2492, 2014.

[Shi *et al.*, 2015] Chuan Shi, Zhiqiang Zhang, Ping Luo, Philip S. Yu, Yading Yue, and Bin Wu. Semantic path based personalized recommendation on weighted heterogeneous information networks. In *CIKM*, pages 453–462. ACM, 2015.

[Singh and Gordon, 2008] Ajit P. Singh and Geoffrey J. Gordon. Relational learning via collective matrix factorization. In *SIGKDD*, pages 650–658. ACM, 2008.

[Sun *et al.*, 2011] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *PVLDB*, 4(11):992–1003, 2011.

[Tang *et al.*, 2012] H. Tang, J.and Gao, H. Liu, and A. Das Sarma. eTrust: Understanding trust evolution in an online world. In *SIGKDD*, pages 253–261. ACM, 2012.

[Vahedian, 2014] Fatemeh Vahedian. Weighted hybrid recommendation for heterogeneous networks. In *RecSys*, pages 429–432. ACM, 2014.

[Wang and Mahadevan, 2011] Chang Wang and Sridhar Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *IJCAI*, pages 1541–1546, 2011.

[Yu *et al.*, 2013] Xiao Yu, Xiang Ren, Quanquan Gu, Yizhou Sun, and Jiawei Han. Collaborative filtering with entity similarity regularization in heterogeneous information networks. In *IJCAI HINA*, 2013.

[Zafarani and Liu, 2013] Reza Zafarani and Huan Liu. Connecting users across social media sites: a behavioral-modeling approach. In *KDD*, pages 41–49, 2013.

[Zhang and Yu, 2015] Jiawei Zhang and Philip S Yu. Community detection for emerging networks. In *SDM*. SIAM, 2015.

[Zhang *et al.*, 2006] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. Learning from incomplete ratings using non-negative matrix factorization. In *SDM*, pages 549–553, 2006.

[Zhang *et al.*, 2015] Jiawei Zhang, Weixiang Shao, Senzhang Wang, Xiangnan Kong, and Philip S Yu. PNA: Partial network alignment with generic stable matching. In *IRI*, pages 166–173. IEEE, 2015.