3-2018

Do your friends make you buy this brand?: Modeling social recommendation with topics and brands

Minh Duc LUU
Singapore Management University, mdluu.2011@phdis.smu.edu.sg

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

Citation

LUU, Minh Duc and LIM, Ee Peng. Do your friends make you buy this brand?: Modeling social recommendation with topics and brands. (2018). *Data Mining and Knowledge Discovery*. 32, (2), 287-319.
Available at: https://ink.library.smu.edu.sg/sis_research/3783

Published in Data Mining and Knowledge Discovery, 2017 September, Pages 1-33
http://doi.org/10.1007/s10618-017-0535-9

Do your friends make you buy this brand?

Modeling social recommendation with topics and brands

Minh-Duc Luu[1] · Ee-Peng Lim[1]

Abstract Consumer behavior and marketing research have shown that brand has significant influence on product reviews and product purchase decisions. However, there is very little work on incorporating brand related factors into product recommender systems. Meanwhile, the similarity in brand preference between a user and other socially connected users also affects her adoption decisions. To integrate seamlessly the individual and social brand related factors into the recommendation process, we propose a novel model called **Social Brand–Item–Topic** (**SocBIT**). As the original SocBIT model does not enforce non-negativity, which poses some difficulty in result interpretation, we also propose a non-negative version, called **SocBIT^{+}**. Both SocBIT and SocBIT^{+} return not only user topic interest, but also brand-related user factors, namely *user brand preference* and *user brand-consciousness*. The former refers to user preference for each brand, the latter refers to the extent to which a user relies on brand to make her adoption decisions. Our experiments on real-world datasets demonstrate that SocBIT and SocBIT^{+} significantly improve rating prediction accuracy over state-of-the-art models such as *Social Regularization* Ma et al. (in: ACM conference on web search and data mining (WSDM), 2011), *Recommendation by Social Trust Ensemble* Ma et al. (in: ACM conference on research and development in information retrieval (SIGIR), 2009a) and *Social Recommendation* Ma et al. (in: ACM conference on information and knowledge management (CIKM), 2008), which incorporate only the social factors. Specifically, both SocBIT and SocBIT$^+$ offer an improvement of at least 22% over these state-of-the-art models in rating prediction for various real-world datasets. Last but not least, our models also outperform the mentioned models in adoption prediction, e.g., they provide higher precision-at-N and recall-at-N.

Keywords Brand effect · Adoption · Social recommendation · Probabilistic matrix factorization · Latent factors

1 Introduction

Our behaviors of adopting items are determined by several personal factors (e.g., interest, budget constraint, and brand preference); social factors (e.g., friends adopting the same items); item factors (e.g., item features and brand) and other external marketing event factors. Modeling how these factors interact with one another as item adoptions occur is an important research problem as these factors can help build more accurate item search and recommendation applications. Ideally, we would like to consider all these factors in a single model. In this work, we address this goal by focusing on modeling item adoptions that can be attributed to brand among other factors.

We first define the concept of brand to distinguish it from other attributes of items. We adapt the following definition of *brand* from American Marketing Association (AMA) dictionary:[1] "A brand is a name, term, design, symbol, or any other feature that identifies one seller's goods or services as distinct from those of other sellers." In many settings, brands are the sellers or creators of the goods/services themselves.We thus offer a simple brand definition, i.e., *a brand is an entity which creates items/services that can be differentiated from the items/services of others*. With this definition, brands can not only be companies, but also individuals that are involved in the creation of products and services. For instance, Steven Spielberg is a "brand" among movie directors as he creates blockbuster movies; Richard Branson is a brand as he is known to found successful businesses.

It is noteworthy that brand is more than just a simple feature of item like color, size etc. A brand encompasses several aspects, it can be a symbol of quality (e.g. prestigious brands) and style (e.g. fashion brands) or can be both. A brand is also usually associated with customer service. A simple meta-information alone cannot represent all of these. Neither can topic of item. Thus, we believe that modelling brand is much different from modelling meta-information of item and/or modelling topic of item.

The presence of brand leads to new challenges in modeling user–item adoptions. Firstly, every user may have her own brand preferences (or awareness) and users decide adopted items considering brands at different degrees. A user may adopt items based on brand preferences only, on topical interest only, or a mixture of both. When a user does not depend on brand, her brand preferences may become unimportant. On the other extreme, another user may adopt items completely based on brand (i.e., a brand conscious user), making the topical interests less important.

The second challenge is brought about by the co-mingling between brand and social connections in adoption decisions. Just as a user's interest topics could be influenced by the interest topics of socially connected users, the user's brand preferences could also be influenced by those of socially connected users. Modeling the brand preferences and brand based adoption decisions in the context of social network is therefore essential.

1.1 Research objectives and contributions

While there have been extensive works on recommender systems, including traditional matrix factorization based recommenders (Koren 2008; Koren et al. 2009) and social recommenders (Jiang et al. 2012; Ma et al. 2008, 2009a, 2011b), they do not explicitly model brand related factors which co-mingle with social factors. To develop recommender systems that can leverage on brands, we need new methods that address the above two challenges. In this paper, we therefore seek to develop new matrix factorization based recommendation models that consider brand-related factors and social factors influencing user–item ratings. We also study how well the new method is, in comparison to those without considering brand-related factors. Ideally, we also want the new method to yield interpretable results which help us learn the dependency of users on making adoption decisions using brand. The novel contributions of this work can be summarized as follows:

– We propose a novel matrix factorization-based model, called **Social Brand–Item–Topic Model (SocBIT)** that incorporates both brand factors and social homophily. Other than modeling each user and item topic factors as in existing models (Jiang et al. 2012; Liu et al. 2015; Ma et al. 2008, 2009a, 2011b), SocBIT goes further by assigning each user and item a set of *brand factors* and learns the *brand-consciousness* level of users. SocBIT models social homophily by facilitating socially connected users to share similarity in topic and brand factors.
– SocBIT does not enforce non-negative constraints on topic factors, which makes it hard to interpret the learnt factors. We therefore propose a non-negative version for SocBIT, called **SocBIT$^+$**. Our experiments show that SocBIT$^+$ yield interpretable topic and brand related factors while achieving prediction accuracy similar to that of SocBIT.
– Our experiments on real-world datasets from Flixster, FourSquare and ACM Digital Library (ACMDL) show that both SocBIT and SocBIT$^+$ significantly improve user–item rating prediction accuracy over state-of-the-art models, namely, SoReg (Ma et al. 2011b), RSTE (Ma et al. 2009a) and SoRec (Ma et al. 2008). The improvement, in terms of Root Mean Square Error, is at least 25.8% on Flixster data, 31.8% on FourSquare data, and 22% on ACMDL data. Our proposed models also achieve better precision-at-N and recall-at-N than the three state-of-the-art models.
– Finally, we are able to derive empirical findings using SocBIT$^+$ so as to answer a few research questions regarding our real datasets, namely: (a) what is the proportion of brand conscious users in the studied user population? (b) how are brand conscious users different from other users? (c) how are the brands preferred by brand conscious users different from those by other users?

The rest of the paper is organized as follows. First, in Sect. 2, we review related work and highlight SoRec, the model that inspired SocBIT. Next, we describe datasets and some empirical analysis motivating assumptions of our models in Sect. 3. We then provide the formulation and inference of SocBIT and SocBIT$^+$ in Sect. 4. We then show our evaluation experiments on real-world datasets in Sect. 5. Finally, we provide conclusion and discussions on future work in Sect. 6.

2 Related work

Our research falls within the domain of matrix factorization-based recommender systems, including the emerging social recommenders. In this section, we therefore review the related works in matrix factorization-based recommender systems and social recommendation methods. We also compare our work with existing brand-based recommendation methods.

2.1 Traditional matrix factorization

Matrix factorization (MF) based recommendation methods have been shown to yield accurate results in recent years and widely adopted by the Information Retrieval (Hofmann 2003, 2004), Machine Learning (Mnih and Salakhutdinov 2007; Salakhutdinov and Mnih 2008), and data mining research communities (Koren 2008; Koren et al. 2009; Baltrunas et al. 2011; Su et al. 2013). The common idea among these methods is to derive user and item latent factors from the user–item rating matrix; and to predict missing ratings using the similarity between these user and item latent factors. In other words, each user u is associated with a vector $\boldsymbol{\theta}_u$ of latent factors and each item i is associated with a vector $\boldsymbol{\theta}_i$ of latent factors. Both vectors share the same dimension K, which is much smaller than the numbers of users and items. For a given user u, the entries in $\boldsymbol{\theta}_u$ measure the interest of u over the K factors, which can be interpreted as topics in certain contexts. Meanwhile, the entries in $\boldsymbol{\theta}_i$ measure the extent to which item i is relevant to the K factors. As a result, the inner product $\boldsymbol{\theta}_u^T \boldsymbol{\theta}_i$ captures how much i matches u's interest, thus can be used to approximate the observed rating $r_{u,i}$ as

$$r_{u,i} \approx \hat{r}_{u,i} = \boldsymbol{\theta}_u^T \boldsymbol{\theta}_i \tag{1}$$

or in matrix form

$$\mathbf{R} \approx \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I \tag{2}$$

where $\mathbf{R} = (r_{u,i})_{u,i}$ is the matrix of observed ratings, $\boldsymbol{\Theta}_U$ and $\boldsymbol{\Theta}_I$ are two low-rank matrices of user and item latent factors respectively. MF methods thus learn the latent factors by minimizing the total regularized squared error:

$$\min_{\boldsymbol{\Theta}_U, \boldsymbol{\Theta}_I} \|\mathbf{R} - \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I\|_F^2 + \lambda \left(\|\boldsymbol{\Theta}_U\|_F^2 + \|\boldsymbol{\Theta}_I\|_F^2 \right) \tag{3}$$

The regularization term weighted by λ is used for penalizing the magnitudes of learned parameters so as to avoid overfitting due to the sparsity of rating matrix. The constant λ

controls the extent of regularization. Once all $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_i$'s are learned, one can predict the rating of any pair (u', i') by the inner product of $\boldsymbol{\theta}_{u'}$ and $\boldsymbol{\theta}_{i'}$. Ruslan Salakhutdinov and Andriy Mnih proposed a probabilistic explanation for the regularization (Mnih and Salakhutdinov 2007; Salakhutdinov and Mnih 2008). They reformulated the minimization problem to its equivalent form of maximizing a posterior where the conditional probability corresponds to the error term and the priors form the regularization term in Eq. (3). We will also use this alternative in our model formulation in Sect. 4.

The non-negative variants of MF (Lee and Seung 1999, 2001; Paatero and Tapper 1994; Xu et al. 2003; Hoyer 2002, 2004; Ding et al. 2006a; Lin 2007) impose non-negativity constraints on $\boldsymbol{\Theta}_U$ and $\boldsymbol{\Theta}_I$ so that the learned factors can be better interpreted, e.g., in terms of topic for document clustering (Xu et al. 2003; Yang et al. 2005; Ding et al. 2005, 2006a, b; Shahnaz et al. 2006). There are also efficient implementations for MF (Lin 2007; Pilászy et al. 2010) or variants which, instead of low-rank factor matrices, look for low-norm ones (Srebro et al. 2004; Rennie and Srebro 2005; Weimer et al. 2007, 2009). All these methods however do not consider other additional factors that may affect the generation of user–item ratings.

2.2 Social recommendation methods

Inspired by the idea that users' ratings of items may be influenced by users' friends, MF approach has been extended to consider social connections among users (Bedi et al. 2007; Massa and Avesani 2007; Ma et al. 2008, 2009a, b, 2011a, b, c; Jamali and Ester 2010). These connections may be friendships, trusts, follow links or others. By incorporating the observed social connection into MF, it has been shown that user latent factors, item latent factors and social influence can be jointly learned. Moreover, user preference and social influence are proved to be complementary factors which boost recommendation accuracy (Jiang et al. 2012).

According to Tang et al. (2013), social recommendation models can be further divided into the following sub-categories:

- Co-factorization models (Ma et al. 2008; Tang et al. 2013): these models assume that each user applies the same latent preference in both assigning ratings and making social relationships. The models thus jointly factorize the rating matrix $\mathbf{R}$ and user–user social weight matrix $\mathbf{W}$ to learn user and item latent factors.
- Ensemble models (Ma et al. 2009a; Tang et al. 2012): the models rely on the idea that users and their friends usually have similar ratings on items, thus a user's missing rating can be estimated as some kind of average of known ratings from the user herself and her friends.
- Social regularization models (Ma et al. 2011b; Jamali and Ester 2010; Jiang et al. 2012): these models assume that a user's preference is similar to that of her friends and thus propose various regularization techniques to ensure the assumption.

We now review a representative of each category. We choose the representatives which are most closely related to our work so that we can compare with them later in experiments.

A representative of *co-factorization* models is a probabilistic MF model called SoRec (Ma et al. 2008). The model proposes the following factorization.

$$\mathbf{R} \approx \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I \quad \text{and } \mathbf{W} \approx \boldsymbol{\Theta}_U^T \mathbf{Z} \tag{4}$$

where, in addition to the user and item latent factor matrices $\boldsymbol{\Theta}_U$ and $\boldsymbol{\Theta}_I$, $\mathbf{Z}$ is another user latent factor matrix for generating the social weight matrix. The model then learns user and item latent factors by solving the following optimization problem:

$$\min_{\boldsymbol{\Theta}_U, \boldsymbol{\Theta}_I, \mathbf{Z}} \|\mathbf{R} - \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I\|_F^2 + \alpha \|\mathbf{W} - \boldsymbol{\Theta}_U^T \mathbf{Z}\|_F^2 + \lambda \left(\|\boldsymbol{\Theta}_U\|_F^2 + \|\boldsymbol{\Theta}_I\|_F^2 + \|\mathbf{Z}\|_F^2 \right) \tag{5}$$

Introducing $\mathbf{Z}$ as a second user latent factor matrix, however, reduces the interpretability of SoRec model.

For *ensemble* models we have the so-called Recommendation by Social Trust Ensemble (RSTE, Ma et al. 2009a). The model proposes to estimate rating of user u on item i as

$$r_{u,i} \approx \boldsymbol{\theta}_u^T \boldsymbol{\theta}_i + \beta \sum_{v \in N_u} s_{u,v} \boldsymbol{\theta}_v^T \boldsymbol{\theta}_i \tag{6}$$

where N_u is the set of friends of u and $s_{u,v}$ is the observed similarity in rating vectors of u and v. The constant β controls the influence of social information on ratings. The matrix form of (6) is

$$\mathbf{R} \approx (\mathbf{I} + \beta \mathbf{S}) \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I \tag{7}$$

where $\mathbf{I}$ denotes the identity matrix. Inference of RSTE thus involves solving the following optimization problem:

$$\min_{\boldsymbol{\Theta}_U, \boldsymbol{\Theta}_I} \|\mathbf{R} - (\mathbf{I} + \beta \mathbf{S}) \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I\|_F^2 \tag{8}$$

Finally, for *social regularization* models, we have the so-called Social Regularization (SoReg, Ma et al. 2011b). The intuition of SoReg is that friends in a user's social network may have diverse tastes. Thus, the model proposes a pair-wise regularization as,

$$\min \sum_u \sum_{v \in N_u} s_{u,v} \|\boldsymbol{\theta}_u - \boldsymbol{\theta}_v\|^2 \tag{9}$$

where $s_{u,v}$ again denotes the similarity based on previous ratings. The similarity can be computed by Pearson Correlation Coefficient or Cosine similarity of commonly rated items by u and v. Adding this regularization with the error term forms the objective function of SoReg:

$$\min_{\boldsymbol{\Theta}_U, \boldsymbol{\Theta}_I} \|\mathbf{R} - \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I\|_F^2 + \alpha \sum_u \sum_{v \in N_u} s_{u,v} \|\boldsymbol{\theta}_u - \boldsymbol{\theta}_v\|^2 + \lambda \left(\|\boldsymbol{\Theta}_U\|_F^2 + \|\boldsymbol{\Theta}_I\|_F^2 \right) \tag{10}$$

On the whole, all these models extend the traditional MF approach by incorporating the social network information. Although none of them consider brand factors, SoReg, RSTE and SoRec are state-of-the-art models and provide important ideas for our approach. Thus, we will later compare our models against the models in our experiment evaluation.

2.3 Brand, consumer decisions and brand-based recommendation

According to Belén del Río et al. (2001), brands are important in item adoption and recommendation since they provide the following functions: (i) guarantee, (ii) personal identification, (iii) social identification, and (iv) status.

The *guarantee* function refers to the ability of brands to provide quality assurance, meeting consumer expectations and reducing perceived risks, especially when a consumer has to choose an item in a unfamiliar topic or under uncertainty (Ambler 1997; Erdem et al. 1999, 2004; Erdem and Keane 1996; Ubilava et al. 2011). The *personal identification* function refers to consumers identifying themselves with certain brand. The greater the consistency between the brand image and the consumer's self-image, the larger is her preference toward the brand and the more likely she adopts items from the brand (Graeff 1996; Hogg et al. 2000; Belén del Río et al. 2001). The *social identification* function refers to the brand's ability to help its consumers' to be either identified in or differentiated from her group of peers [Optimal Distinctiveness theory (Brewer 1991; Long and Schiffman 2000)]. Finally, the *status* function refers to the admiration and prestige a consumer may enjoy by adopting items from a brand (Solomon 1999; Vigneron and Johnson 1999).

Although brand has such important functions in user–item adoption, brand-based recommendation receives much less attention compared with previous approaches. There are very few works focusing on modeling brand effect on ratings and item adoptions (Zhang and Pennacchiotti 2013; Wakita et al. 2015). In Zhang and Pennacchiotti (2013), the authors studied the correlation between the brands "liked" by a social media user and his items purchased to make recommendations of items of new brands. The work focuses on user brand preference but overlooks the social network information. It also does not study the level user depends on brand to make purchase decisions. The work (Wakita et al. 2015) proposes to treat fashion brands at the item level and apply recommendation methods on brands directly. This work therefore does not consider recommending specific items nor modeling brand-related factors and social factors.

3 Datasets and empirical analysis

Before we describe our proposed models Social Brand Item Topic (SocBIT) and its non-negative version SocBIT$^+$, we first define the observed rating and social network data to be modeled. We then describe two empirical analysis on real-world data motivating the assumptions of our models.

To investigate the brand effect on item adoptions, we have gathered three real world datasets from Flixter, FourSquare and ACM Digital Library (ACMDL). The same datasets will also be used in our subsequent experiments (see Sect. 5). Detailed statistics of the datasets are provided in Table 1.

Table 1 Statistics of datasets

Dataset	# Users (N)	# Items (M)	# Brands (Q)	# Ratings	# Edges
FL-IMDB	142,162	25,242	22,879	6,643,917	2,289,524
4SQDB	4940	14,821	6237	101,680	85,188
ACMDB	163,511	299,724	22,294	1,577,948	922,979

3.1 Datasets

3.1.1 Dataset from Flixster and IMDB (FL-IMDB)

Flixster is a social movie site where users share movie reviews and ratings. In this dataset, movies and directors of movies are regarded as items and brands respectively. Considering directors as movie brands is reasonable because they play major role in creating movies and they can affect movie rating/adoption, especially if the director(s) are popular. For example, movies directed by Steven Spielberg usually attract more audience than those directed by a normal director. The ratings are 10 discrete values in the range [0.5,5] with step size 0.5. We scale the ratings to range [0,1] using the transformation $f(x) = (x - 0.5)/4.5$. Users also have undirected friend relationships. We obtained the original Flixster dataset publicly available from Jamali (2010). The dataset contains ratings of movies from December 1941 to November 2009. However, it does not contain information of movie directors, the needed information of movie brands.

To obtain such brand information, we join the Flixter dataset with a dataset of director-movie relationships we crawled from IMDB. The join is based on exact match on the movie name and the year. This gives us the final dataset denoted as **FL-IMDB**. Note that **FL-IMDB** is a subset of the original Flixster data. Its statistics are given in Table 1.

3.1.2 Dataset from FourSquare (4SQDB)

FourSquare is a popular location-based social network (LBSN) which allows users and venues to interact with one another. In this dataset, we consider restaurant venues as items and the chains they belong to as brands (e.g. KFC). If the venue does not belong to any chain then we consider the venue itself as a brand.Users can follow other users. The follow network is represented as a directed graph $G = (U, \mathbf{W})$, where the weights in $\mathbf{W}$ are binary: $w_{u,v} = 1$ if v follows u, and 0 otherwise. Here users do not rate venues. Instead, they can perform *check-ins* on venues. Users can also write tips (a kind of short review) on venues. When a user u checks in at a venue i, u is said to adopt i, i.e., the rating $r_{u,i} = 1$.

We collected raw check-in data on FourSquare from June 2011 to October 2015 via tweets of FourSquare users in Singapore. These check-in tweets were determined by their latitude and longitude. In our experiments, we focus only on check-ins to *food venues* because of two reasons: (i) food venues contribute the largest number of check-ins compared to other kinds of venues, and (ii) focusing on only one type of venues lead to more interpretable results, especially for the subsequent evaluation of brand-conscious users. We also filtered out low-activity users with less than 3 check-ins and venues adopted by these users. After this filtering, we are left with 4940 users, each of them adopted at least 3 of 14,821 food venues.

Adoption data just provides us with ratings of value 1. We thus need to impute 0-ratings into the rating matrix **R**. For that, we first divide all venues into 50 m × 50 m grid cells. For each pair (u, i) with $r_{u,i} = 1$, we randomly sample a venue j from other venues in the same cell such that u has not checked in and assign $r_{u,j} = 0$. All remaining $r_{u,i}$'s which are neither 1 nor 0 are assigned "undefined". In other words, we assume that a user checking into a venue i but not the nearby venue j has no interest in j. After this, we obtain the final dataset in standard representation $\mathscr{D} = (\mathbf{R}, G, \mathbf{B})$, which we denote as **4SQDB**.

3.1.3 Dataset from ACM Digital Library (ACMDB)

From ACM Digital Library (ACMDL), we extract data of citations from 1998 to 2010. Each publication record consists of (i) its Id, title and abstract; (ii) its authors; and (iii) reference records; each of which contains Id, title and authors of a cited paper. Each citation is considered as an *adoption* where the cited paper is the adopted *item*. We then consider the *first author* of the citing paper as the *user* who adopts the item. The social network among the users is the weighted co-author network G where the weight on each edge (u, v) is the number of papers co-authored by u and v normalized by their total number of papers.

As the total number of authors, around 350 K, is huge, it is not feasible to consider all of them as brands. Instead, we empirically select those authors with citation count significantly higher than that of a normal author. We plot the distribution of citation count in original data and find that (i) the median number of citations is 3, and (ii) about 25% of authors have at least 9 citations, which is 3 times more than the median. Thus, we decide that only authors with at least 9 citations should be considered as brands. We thus retain only those authors as *brands* and extract (i) their items to form the item set I and the brand–create–item matrix **B**, and (ii) users who adopt at least one of these items to form the set of users U. The co-author network among the users is thus a sub network of the original G, we however still denote it as G for simplicity.

Again, each adoption is a rating with value 1. We then impute 0s to obtain the rating matrix **R** by the proximity-based heuristics as what we did on FourSquare data. However, the similarity between two papers is now defined by the Jaccard similarity between two keyword sets extracted from their abstracts. After this, we obtain the final dataset in standard representation $\mathscr{D} = (\mathbf{R}, G, \mathbf{B})$, which we denote as **ACMDB**.

3.2 Empirical analysis

We now perform two kinds of empirical analysis on **ACMDB**. The first investigates into the existence of brand effect on user–item adoption. We would like to validate the assumption of brand affecting the user adoption decisions. Specifically, we want to demonstrate that brand(s) of an item can affect the number of the item's adopters, depending on ranking(s) of the brand(s). For this, we need an empirical ranking of brands which have no or at least low correlation with the number of adopters. As we have not found such ranking on **4SQDB** or **FL-IMDB**, we resort to only analysis on **ACMDB**.

The second analysis examines the correlation between social tie strength and user–user brand similarity. This correlation study will help to validate the assumption that social tie also plays a role in users' brand choices.

3.2.1 Analysis of brand effect

As we do not have a direct measure for brand value of an author, we use the brand value of the university where the author works as a proxy. We would like to check if authors from high-rank universities *attract more attention* than those from low-rank universities, controlling for the author's research topics, citation counts and country of the university affiliation. We measure the amount of attention received by an author by the number of adopters of her papers. We thus create a dataset of citation counts and adopter counts of authors from US universities by combining various data sources as follows.

- We extracted the citation count and adopter count of authors from ACMDL. An author's citation count and adopter count refer to the number of papers and number of authors citing the papers of the author respectively.
- Author affiliation and topic data were obtained from the author profiles in Google Scholar. As the number of authors we could crawl is limited to about 550 per machine, we only crawled the profiles of authors under two research topics, namely, *data mining* and *distributed systems*. These profiles are found by querying Google Scholar with appropriate query terms and extracting the required fields from the returned author profile results. To query authors under data mining, we used the terms "data mining", "text mining", and "social network mining". To query authors under distributed systems, we used the terms "distributed systems", "concurrent", and "parallel". We also performed a manual check on retrieved authors to remove some exceptions outside the two topics.
- University ranking in the Computer Science discipline was crawled from the website www.topuniversities.com. We obtained from the website ranks of 90 US universities. Empirically, we consider universities with ranks 1–45 (46–90) as high (low) rank universities.

For each topic, we then assign authors into bins by their number of citations such that each bin has 40–50 authors. Figure 1 depicts the adopter counts of authors of both high rank and low rank universities in different citation count bins. To keep the figure simple, we only show five citation count bins in each chart. By analyzing adopter counts of authors in each bin, we found that authors from high rank universities indeed have more adopters than those from low rank universities (see Fig. 1a, b). This difference can be found for the authors under both "data mining" and "distributed systems" topics. The difference is smaller (e.g., around 100 for the (600,700] bin for the "distributed systems" topic) for authors with smaller citation counts but larger (e.g., around 300

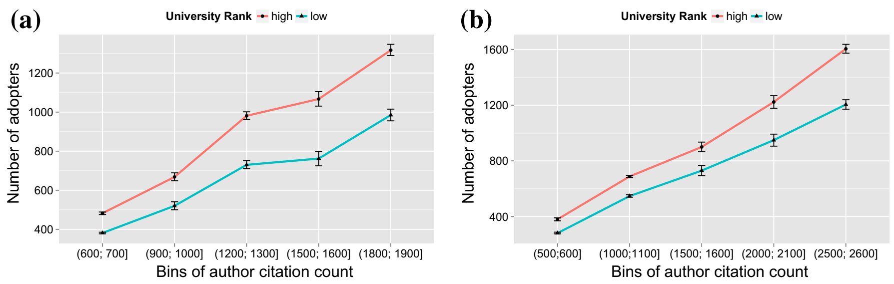


Fig. 1 High-rank-university versus low-rank-university cited authors: a comparison on adopter count. **a** Authors under topic "distributed systems". **b** Authors under topic "data mining"

for the (1800,1900] bin for the "distributed systems" topic) for authors with larger citation counts. This confirms the existence of brand effect.

3.2.2 Analysis of social correlation

Users who are alike choose to be connected with one another, and socially connected users influence one another to be even more similar. These two processes in social networks, known as *selection* and *social influence*, are expected to affect connected users' topic preferences and brand preferences (McPherson et al. 2001; Friedkin 2006). Research has shown that recommendation methods modeling social network effect on users' topic preferences can achieve better recommendation accuracy (Crandall et al. 2008; Ma et al. 2008, 2009a; Jamali and Ester 2010). Inspired by these results, we analyzed ACMDB to confirm a positive correlation between *social tie weight* and user similarity in brand preference, which we call *user–user brand similarity*.

Firstly, we computed the social tie weight $socWeight(u, v)$ as the number of papers co-authored by u and v normalized by their total number of papers:

$$socWeight(u, v) = \frac{|P(u) \cap P(v)|}{|P(u) \cup P(v)|}$$

where $P(x)$ are the papers of which x is one of the authors. We then extracted user pairs who adopted at least one common "brand", i.e., cited at least one common author, and calculated brand similarity for each pair based on authors adopted by both users:

$$brandSim(u, v) = \frac{|A(u) \cap A(v)|}{|A(u) \cup A(v)|}$$

where $A(x)$ are the authors cited by x.

Finally, we computed the Pearson correlation between the user–user brand similarity and the social tie weight. Computing this correlation on the whole co-author network is costly as the network is large with nearly 200,000 users and millions of edges. Thus, we resorted to computing the correlation on samples of randomly chosen users, each of size 1000 users. We iterated this process for 400 samples and plotted

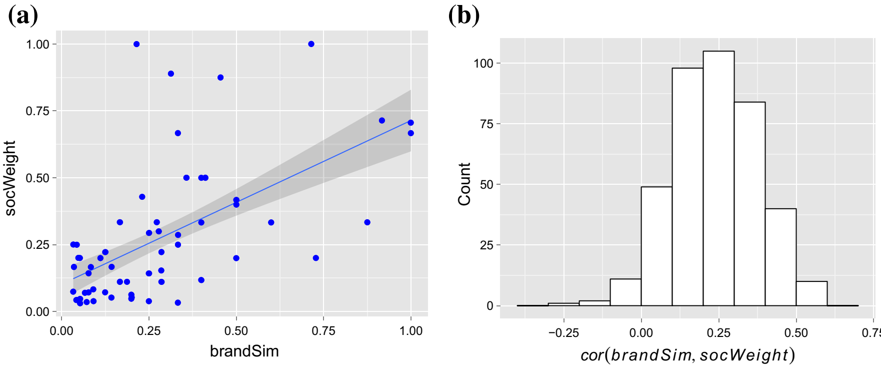


Fig. 2 Correlation between social tie weight and brand-based similarity, observed on citation data from ACMDL. **a** *Scatter plot* with regression line for a sample of 1000 users with 118 pairs of co-authors having brand similarity (*each point* is one pair). The sample's correlation value is 0.5. **b** Empirical distribution of $cor(socWeight, brandSim)$, computed on 400 samples, each consists of 1000 users

the histogram of the correlation values in Fig. 2b. The figure shows that most correlations are positive and their mean is 0.23. This suggests a positive correlation between user–user brand similarity and social tie weight.

4 Proposed concepts and models

Before delving into our proposed models SocBIT and SocBIT$^+$, we first introduce a few important concepts and notations in the subsequent section.

4.1 Concepts and notations

4.1.1 Concepts

Let B denote the set of all brands in data. We first formally define user brand preference as a vector of numeric factors, one for each brand. The factor of each brand indicates how much a user prefers the brand. In the case of SocBIT$^+$, each user–brand factor is non-negative and the larger it is, the more the user prefers the brand.

Definition 1 *(Brand factors of user)* Given a user u, the brand factor $\beta_{b,u}$ of u for a brand $b \in \mathsf{B}$ measures u's preference toward brand b. In the context of SocBIT$^+$, we have $\beta_{b,u} \geq 0, \forall b$.

Next, we define brand factors of item. This arises from the remark that different items of the same brand represent the brand differently, which affects the ratings they receive from users. For example, (i) only signature dishes of a restaurant are the most representative ones and thus preferred by its customers over normal dishes; (ii) among movies of Jackie Chan, the more representative ones receive higher ratings from audience. In other words, for each brand b and each item i that the brand creates, there should be a latent factor measuring the extent to which i represents b. We integrate this into our models by defining this factor as the *brand factor* of item i with respect to brand b. When an item is not created by a brand, the corresponding item–brand factor is 0. In short, we associate each item with a vector of item–brand factors as follows.

Definition 2 *(Brand factors of item)* Given an item i and the set B_i of brands creating i, the brand factor $\beta_{b,i}$ of item i for brand $b \in \mathsf{B}_i$ measures the extent to which i represents b. For $b \notin \mathsf{B}_i$, i.e. brand b does not create item i, $\beta_{b,i}$ is simply 0. In the context of SocBIT$^+$, we have $\beta_{b,i} \geq 0, \forall b \in \mathsf{B}$.

For a user u to adopt an item i based on brand, u and i should have high user–item brand-based similarity, which can be measured by the sum $\sum_{b \in \mathsf{B}_i} \beta_{u,b}\beta_{i,b}$. When u adopts an i based on topic, the user–item topic similarity will be used as in matrix factorization models.

Assumption 1 *(Topic-based and Brand-based Ratings)* The rating a user gives to an item can be approximated as a weighted average of topic-based and to brand-based similarities. The weight is user dependent.

Our analysis of social correlation in Sect. 3.2.2 suggests that user–user brand similarity is higher for users with stronger social ties. By combining the correlation of topics and brands among socially connected users, we propose the second assumption.

Assumption 2 *(Social Correlation)* Social tie strength between any two users **correlates** with their topic-based and brand-based similarities.

In this paper, we use the following notations.

4.1.2 Notations

Given a set of users $U = \{u_1, \ldots, u_N\}$ and a set of items $I = \{i_1, \ldots, i_M\}$, the ratings of the users on the items are represented by a $N \times M$ user–item matrix $\mathbf{R} = (\{r_{u,i}\})$. A rating $r_{u,i}$ is undefined when u has not rated i. Otherwise, $r_{u,i}$ can be any real numbers in [0, 1] after normalization. The users U are connected by a (un)directed social network $G = (U, \mathbf{W})$ where $\mathbf{W}$ is the $N \times N$ matrix of non-negative edge weights. A positive weight $w_{u,v}$ represents the strength of social influence between users u and v, while a zero weight $w_{u,v}$ represents no connection. We require the weights to be in [0, 1]. We represent the brand–create–item relationship by a $Q \times M$ matrix $\mathbf{B}$ with binary values. $b_{i,j} = 1$ when item i belongs to brand j, and 0 otherwise. In short, the standard representation of a dataset in this work is $\mathscr{D} = (\mathbf{R}, G, \mathbf{B})$. All these notations are summarized in Table 2.

We are now ready to formulate SocBIT model based on matrix factorization framework. We first describe how SocBIT jointly models the generative processes of ratings and social weights. We then propose SocBIT's conditional probabilities based on the generative processes. Finally, we derive SocBIT's posterior from the conditional probabilities and Gaussian priors.

Table 2 Notations used in this paper

Symbol	Description
U ($\lvert U\rvert = N$)	Set of users
I ($\lvert I\rvert = M$)	Set of items
Q	Number of brands
K	Number of latent topics
$\mathbf{R} = (r_{u,i})_{U,I}$	Rating matrix
$G = (U, \mathbf{W})$	Directed, weighted network among users
$\mathbf{B}$	Matrix of brand–create–item relationships
$\mathscr{D} = (\mathbf{R}, G, \mathbf{B})$	Standard representation of a dataset
$\boldsymbol{\theta}_u$ and $\boldsymbol{\beta}_u$	Topic and brand factors of user u
$\boldsymbol{\theta}_i$ and $\boldsymbol{\beta}_i$	Topic and brand factors of item i
$w^{\mathbf{t}}_{u,v}$, $w^{\mathbf{b}}_{u,v}$ and $w_{u,v}$	Topic-based, brand-based and total influence of u on v
$r^{\mathbf{t}}_{u,i}$, $r^{\mathbf{b}}_{u,i}$ and $r_{u,i}$	Topic-based, brand-based and total rating of u for i
δ_u	Topic dependency weight of u
$\boldsymbol{\delta}_U = (\delta_u)_{u\in U}$	*Topic dependency* vector of users

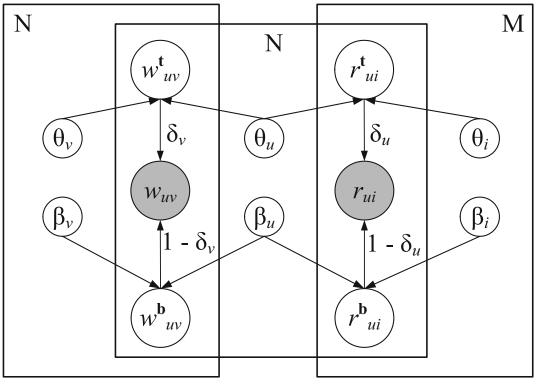
4.2 Generative process

The plate diagram of SocBIT is given in Fig. 3. The left plate represents users and how social connections among them are generated. The right plate represents items and how ratings are generated. Similar to traditional MF, each user u and item i is assigned a user topic vector $\boldsymbol{\theta}_u$ and item topic vector $\boldsymbol{\theta}_i$ respectively. Both vectors are K dimensional w.r.t. K topics. By Definition 1, SocBIT models brand preference of user u by a vector $\boldsymbol{\beta}_u = (\beta_{b,u})_{b\in\mathsf{B}}$. By Definition 2, item i is associated with a vector $\boldsymbol{\beta}_i = (\beta_{b,i})_{b\in\mathsf{B}}$, where each element $\beta_{b,i}$ measures how much i represents brand b. Both $\boldsymbol{\beta}_u$ and $\boldsymbol{\beta}_i$ vectors are Q-dimensional corresponding to the number of brands. Finally, when i does not belong to brand b, i.e., the entry $B_{i,b}$ in brand–create–item matrix $\mathbf{B}$ is 0, the factor $\boldsymbol{\beta}_{b,i}$ is also 0.

SocBIT models the generation process of each rating $r_{u,i}$ as follows. First, the topic vectors of u and i are used to generate a **topic-based** rating $r^{\mathbf{t}}_{u,i}$. Meanwhile, the brand vectors of u and i are used to generate a **brand-based** rating $r^{\mathbf{b}}_{u,i}$. By Assumption 1, the final rating $r_{u,i}$ is then approximated as a weighted average of $r^{\mathbf{t}}_{u,i}$ and $r^{\mathbf{b}}_{u,i}$. The weights depend on how much user u depends on topics to assign ratings, which is measured by the *topic dependency weight* $\delta_u \in [0, 1]$. In short, by defining the ratings $r^{\mathbf{t}}_{u,i}$, $r^{\mathbf{b}}_{u,i}$ as

$$r^{\mathbf{t}}_{u,i} = \boldsymbol{\theta}_u^T \boldsymbol{\theta}_i, \quad r^{\mathbf{b}}_{u,i} = \boldsymbol{\beta}_u^T \boldsymbol{\beta}_i$$

Fig. 3 Graphical model for SocBIT



we can approximate $r_{u,i}$ by

$$r_{u,i} \approx \widehat{r}_{u,i} \stackrel{\text{def}}{=} g\left(\delta_u r^{\mathbf{t}}_{u,i} + (1-\delta_u) r^{\mathbf{b}}_{u,i}\right) = g\left(\overbrace{\delta_u \boldsymbol{\theta}_u^T \boldsymbol{\theta}_i + (1-\delta_u)\boldsymbol{\beta}_u^T \boldsymbol{\beta}_i}^{\gamma_{u,i}}\right) = g(\gamma_{u,i}) \tag{11}$$

where the logistic function $g(x) = 1/(1+e^{-x})$ is used to bound the approximated rating in [0, 1].

Other than observed ratings, SocBIT also models observed social network connections $\mathbf{W}$. The left plate in Fig. 3 denotes the users u has influence over. SocBIT assumes the social weight u exerts on v is accounted by two similarities between them. The first is topic-based similarity $w^{\mathbf{t}}_{u,v}$ and the second is brand-based similarity $w^{\mathbf{b}}_{u,v}$. The final social weight is thus a weighted average of $w^{\mathbf{t}}_{u,v}$ and $w^{\mathbf{b}}_{u,v}$ as follows.

$$w_{u,v} \approx \widehat{w}_{u,v} \stackrel{\text{def}}{=} g\left(\overbrace{\delta_v \boldsymbol{\theta}_u^T \boldsymbol{\theta}_v + (1-\delta_v)\boldsymbol{\beta}_u^T \boldsymbol{\beta}_v}^{\omega_{u,v}}\right) = g(\omega_{u,v}) \tag{12}$$

Note that in Eq. 12, δ_v is used instead of δ_u. It is because the weight $w_{u,v}$ measures how much influence that v perceives/receives from u. Thus, it should depend on the extent v relies on topic or brand similarity.

4.3 Conditional probabilities

In this section, we propose the conditional probabilities for rating and social network matrices. For that, we first need to rewrite Eqs. (11) and (12) in matrix factorization form. As the form show how the two matrices are estimated given parameters of SocBIT, it is then straightforward to obtain the conditional probabilities. We need some more matrix notations.

- $\boldsymbol{\Theta}_U = (\boldsymbol{\theta}_u)_U \in \mathbb{R}^{K \times N}$ and $\boldsymbol{\Theta}_I \in \mathbb{R}^{K \times M}$: matrices of topic factors of users and items,
- $\mathscr{B}_U = (\boldsymbol{\beta}_u)_U \in \mathbb{R}^{Q \times N}$ and $\mathscr{B}_I = (\boldsymbol{\beta}_i)_I \in \mathbb{R}^{Q \times M}$: matrices of brand factors of users and items,
- $\boldsymbol{\Delta}_U = diag(\delta_u)_U \in \mathbb{R}^{N \times N}$: diagonal matrix of which diagonal entries are topic dependency weights of users,
- $\pi = (\boldsymbol{\Theta}_U, \boldsymbol{\Theta}_I, \mathscr{B}_U, \mathscr{B}_I, \boldsymbol{\delta}_U)$: all the parameters of SocBIT.

Conditional probability for ratings We now can rewrite Eq. (11) as

$$\mathbf{R} \approx \widehat{\mathbf{R}} \stackrel{\text{def}}{=} g\left(\boldsymbol{\Delta}_U \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_I + (\mathbf{Id} - \boldsymbol{\Delta}_U)\, \mathscr{B}_U^T \mathscr{B}_I\right) \tag{13}$$

where $g(\mathbf{A})$ of a matrix $\mathbf{A}$ is simply the matrix obtained by applying the logistic function on $\mathbf{A}$ element-wise and $\mathbf{Id}$ denotes the identity matrix. This form inspires the conditional probability

$$\begin{aligned} p(\mathbf{R}|\pi;\sigma_R) &= \prod_{(u,i)} \mathscr{N}\left(r_{u,i}|\widehat{r}_{u,i}(\underbrace{\delta_u, \boldsymbol{\theta}_u, \boldsymbol{\theta}_i, \boldsymbol{\beta}_u, \boldsymbol{\beta}_i}_{\pi_{u,i}}); \sigma_R^2\right)^{\mathbb{1}_{u,i}^R} \\ &= \prod_{(u,i)} \mathscr{N}\left(r_{u,i}|\widehat{r}_{u,i}(\pi_{u,i});\ \sigma_R^2\right)^{\mathbb{1}_{u,i}^R} \end{aligned} \tag{14}$$

where $\mathbb{1}_{u,i}^R$ is the indicator on whether u actually rates i.

One may argue that the decomposition in Eq. (13) may be problematic as matrices $\mathscr{B}_U$ and $\mathscr{B}_I$ have high dimension Q. However, in reality, these matrices have lots of 0 entries as each user is only interested in a few brands and each item only belongs to a few brands. Thus, the product $\mathscr{B}_U^T \mathscr{B}_I$ is still equivalent to a low rank factorization.

Conditional probability for social weights Similarly, Eq. (12) can be rewritten as follows.

$$\mathbf{W} \approx \widehat{\mathbf{W}} \stackrel{\text{def}}{=} g\left(\boldsymbol{\Delta}_U \boldsymbol{\Theta}_U^T \boldsymbol{\Theta}_U + (\mathbf{Id} - \boldsymbol{\Delta}_U)\mathscr{B}_U^T \mathscr{B}_U\right) \tag{15}$$

The conditional probability for social matrix $\mathbf{W}$ is then:

$$\begin{aligned} p(\mathbf{W}|\pi, \sigma_W) &= \prod_{u,v} \mathscr{N}\left(w_{u,v}|\widehat{w}_{u,v}(\underbrace{\delta_v, \boldsymbol{\theta}_u, \boldsymbol{\theta}_v, \boldsymbol{\beta}_u, \boldsymbol{\beta}_v}_{\pi_{u,v}}); \sigma_W^2\right)^{\mathbb{1}_{u,v}^W} \\ &= \prod_{u,v} \mathscr{N}\left(w_{u,v}|\widehat{w}_{u,v}(\pi_{u,v}); \sigma_W^2\right)^{\mathbb{1}_{u,v}^W} \end{aligned} \tag{16}$$

where $\mathbb{1}_{u,v}^W$ is the indicator on whether there is a social tie from u to v.

4.4 Posterior of SocBIT

Similar to SoRec (Ma et al. 2008), we use spherical Gaussian distribution as priors. In the following, the standard deviations of the Gaussian distributions for *user–topic, user–brand, item–topic, item–brand* factors and *topic dependency weights* are denoted as $\sigma_{\mathbf{ut}}, \sigma_{\mathbf{ub}}, \sigma_{\mathbf{it}}, \sigma_{\mathbf{ib}}$ and $\sigma_{\mathbf{d}}$ respectively.

– (Priors for user factors)

$$p(\boldsymbol{\Theta}_U|\sigma_{\mathbf{ut}}) = \prod_{u\in U} \mathscr{N}\left(\boldsymbol{\theta}_u|\mathbf{0}_K; \sigma^2_{\mathbf{ut}}\mathbf{Id}\right) \text{ and } p(\mathscr{B}_U|\sigma_{\mathbf{ub}}) = \prod_{u\in U} \mathscr{N}\left(\boldsymbol{\beta}_u|\mathbf{0}_Q; \sigma^2_{\mathbf{ub}}\mathbf{Id}\right) \tag{17}$$

– (Priors for item factors)

$$p\left(\boldsymbol{\Theta}_I|\sigma_{\mathbf{it}}\right) = \prod_{i\in I} \mathscr{N}\left(\boldsymbol{\theta}_i|\mathbf{0}_K; \sigma^2_{\mathbf{it}}\mathbf{Id}\right) \text{ and } p\left(\mathscr{B}_I|\sigma_{\mathbf{ib}}\right) = \prod_{i\in I} \mathscr{N}\left(\boldsymbol{\beta}_i|\mathbf{0}_Q; \sigma^2_{\mathbf{ib}}\mathbf{Id}\right) \tag{18}$$

– (Prior for topic dependency weights) We assume each δ_u has mean 0.5 as most people are neutral, they are neither brand-conscious nor non-brand-conscious.

$$p(\boldsymbol{\delta}_U|\sigma_{\mathbf{d}}) = \prod_{u\in U} \mathscr{N}\left(\delta_u|0.5, \sigma^2_{\mathbf{d}}\right) \tag{19}$$

Given these conditional probabilities and priors, the joint probability is then obtained as

$$\begin{aligned} P(\mathbf{W}, \mathbf{R}, \pi|\Sigma) &= [p(\mathbf{W}|\pi; \sigma_W)p(\mathbf{R}|\pi; \sigma_R)] \\ &\times [p(\boldsymbol{\Theta}_U|\sigma_{\mathbf{ut}})p(\mathscr{B}_U|\sigma_{\mathbf{ub}})p(\boldsymbol{\Theta}_I|\sigma_{\mathbf{it}})p(\mathscr{B}_I|\sigma_{\mathbf{ib}})p(\boldsymbol{\delta}_U|\sigma_{\mathbf{d}})] \end{aligned} \tag{20}$$

where $\Sigma = (\sigma_W, \sigma_R, \sigma_{\mathbf{ut}}, \sigma_{\mathbf{ub}}, \sigma_{\mathbf{it}}, \sigma_{\mathbf{it}}, \sigma_{\mathbf{d}})$ represents all hyper-parameters.

By Bayes theorem, SocBIT's posterior is proportional to its joint probability. Thus, we obtain the negative log posterior as follows.

$$\begin{aligned} -\ln P(\pi|\mathbf{W}, \mathbf{R}; \Sigma) &\propto \frac{1}{\sigma_R^2} \sum_{(u,i)} \mathbb{1}^R_{u,i}\left[\widehat{r}_{u,i}(\pi_{u,i}) - r_{u,i}\right]^2 \\ &+ \frac{1}{\sigma_W^2} \sum_{(u,v)} \mathbb{1}^W_{u,v}\left[\widehat{w}_{u,v}(\pi_{u,v}) - w_{u,v}\right]^2 \\ &+ \frac{1}{\sigma^2_{\mathbf{ut}}}\|\boldsymbol{\Theta}_u\|_F^2 + \frac{1}{\sigma^2_{\mathbf{ub}}}\|\mathscr{B}_u\|_F^2 + \frac{1}{\sigma^2_{\mathbf{ut}}}\|\boldsymbol{\Theta}_i\|_F^2 + \frac{1}{\sigma^2_{\mathbf{ib}}}\|\mathscr{B}_i\|_F^2 + \frac{1}{\sigma^2_{\mathbf{d}}}\|\boldsymbol{\delta}_U - 0.5\|^2 \end{aligned} \tag{21}$$

where $\|.\|_F$ denotes the usual Frobenius norm of matrix.

4.5 SocBIT inference

Maximizing the log posterior over model parameters is equivalent to minimizing the following squared-error objective function with quadratic regularization terms.

$$\mathscr{L}(\pi) = \frac{1}{2}\sum_{(u,i)} \mathbb{1}^R_{u,i}\left[\widehat{r}_{u,i}(\pi_{u,i}) - r_{u,i}\right]^2 + \frac{\lambda_W}{2}\sum_{(u,v)}\left[\widehat{w}_{u,v}(\pi_{u,v}) - w_{u,v}\right]^2$$
$$+ \frac{\lambda^{\mathbf{t}}_U}{2}\|\boldsymbol{\Theta}_U\|^2_F + \frac{\lambda^{\mathbf{b}}_U}{2}\|\mathscr{B}_U\|^2_F + \frac{\lambda^{\mathbf{t}}_I}{2}\|\boldsymbol{\Theta}_I\|^2_F + \frac{\lambda^{\mathbf{b}}_I}{2}\|\mathscr{B}_I\|^2_F + \frac{\lambda_{\mathbf{d}}}{2}\|\boldsymbol{\delta}_U - 0.5\|^2 \tag{22}$$

where regularization coefficients are

$$\left\{\lambda_W, \lambda_{\mathbf{d}}, \lambda^{\mathbf{t}}_U, \lambda^{\mathbf{b}}_U, \lambda^{\mathbf{t}}_I, \lambda^{\mathbf{b}}_I\right\} = \sigma^2_R\left\{1/\sigma^2_W, 1/\sigma^2_{\mathbf{d}}, 1/\sigma^2_{\mathbf{ut}}, 1/\sigma^2_{\mathbf{ub}}, 1/\sigma^2_{\mathbf{it}}, 1/\sigma^2_{\mathbf{ib}}\right\}.$$

To reduce model complexity and avoid overfitting, we set $\lambda^{\mathbf{t}}_U = \lambda^{\mathbf{t}}_I = \lambda_{\mathbf{t}}$ and $\lambda^{\mathbf{b}}_U = \lambda^{\mathbf{b}}_I = \lambda_{\mathbf{b}}$.

A local minimum of this objective function can be found by performing projected gradient descent on the model parameters. The projection is needed to ensure that $\beta_{b,i} = 0$ when $B_{b,i} = 0$. We now show formulae of the gradients.

Gradients for item factors When we derive the gradients of objective function for a given item i, the second term will vanish as it does not involve items. Thus, each of the gradients w.r.t. $\boldsymbol{\theta}_i$ and $\boldsymbol{\beta}_i$ depends on only two components: (i) rating estimations, and (ii) regularizers. We thus have

$$\nabla_{\boldsymbol{\theta}_i}\mathscr{L} = \lambda^{\mathbf{t}}_I\boldsymbol{\theta}_i + \sum_{u\in U}\delta_u\mathbb{1}^R_{u,i}(\widehat{r}_{u,i} - r_{u,i})g'(\gamma_{u,i})\boldsymbol{\theta}_u \tag{23}$$

and

$$\nabla_{\boldsymbol{\beta}_i}\mathscr{L} = \lambda^{\mathbf{b}}_I\boldsymbol{\beta}_i + \sum_{u\in U}(1-\delta_u)\mathbb{1}^R_{u,i}(\widehat{r}_{u,i} - r_{u,i})g'(\gamma_{u,i})\boldsymbol{\beta}_u \tag{24}$$

Gradients for user factors For a given user u, each of the gradients w.r.t $\boldsymbol{\theta}_u$ and $\boldsymbol{\beta}_u$ depends on three components: (i) rating estimations, (ii) weight estimations, and (iii) regularizers. Thus, we have

$$\nabla_{\boldsymbol{\theta}_u}\mathscr{L} = \lambda^{\mathbf{t}}_U\boldsymbol{\theta}_u + \delta_u$$
$$\left[\sum_{i\in I}\mathbb{1}^R_{u,i}(\widehat{r}_{u,i} - r_{u,i})g'(\gamma_{u,i})\boldsymbol{\theta}_i + \lambda_W\sum_{v\in U}\mathbb{1}^W_{u,v}(\widehat{w}_{u,v} - w_{u,v})g'(\omega_{u,v})\boldsymbol{\theta}_v\right] \tag{25}$$

and

$$\nabla_{\boldsymbol{\beta}_u}\mathscr{L} = \lambda_U^{\mathbf{b}}\boldsymbol{\beta}_u + (1-\delta_u)$$
$$\left[\sum_{i\in I}\mathbb{1}_{u,i}^{R}(\widehat{r}_{u,i} - r_{u,i})g'(\gamma_{u,i})\boldsymbol{\beta}_i + \lambda_W\sum_{v\in U}\mathbb{1}_{u,v}^{W}(\widehat{w}_{u,v} - w_{u,v})g'(\omega_{u,v})\boldsymbol{\beta}_v\right] \quad (26)$$

Derivatives for topic dependency weights For a given user u, the derivative of the objective function for topic dependency weight δ_u is:

$$\frac{\partial\mathscr{L}}{\partial\delta_u} = \lambda_{\mathbf{d}}(\delta_u - 0.5) + \sum_{i\in I}\mathbb{1}_{u,i}^{R}(\widehat{r}_{u,i} - r_{u,i})\left(\boldsymbol{\theta}_u^T\boldsymbol{\theta}_i - \boldsymbol{\beta}_u^T\boldsymbol{\beta}_i\right)g'(\gamma_{u,i})$$
$$- \lambda_W\sum_{v\in U}\mathbb{1}_{u,v}^{W}(\widehat{w}_{u,v} - w_{u,v})\left(\boldsymbol{\theta}_u^T\boldsymbol{\theta}_v - \boldsymbol{\beta}_u^T\boldsymbol{\beta}_v\right)g'(\omega_{u,v}) \quad (27)$$

4.6 Nonnegative version: SocBIT$^+$

One issue of the GD inference is that its *additive* update rules cannot guarantee the non-negativity of user and item factors. Using the approach in Lee and Seung (2001), we replace additive update rules by *multiplicative* update rules, which grants us the desired non-negativity. The core idea of the approach is choosing suitable step size to cancel out negative parts in update formulae as follows.

We start with the gradient for updating item topic factors in Eq. (23). The corresponding update formula, in element-wise form, is then

$$\theta_{i,k} \leftarrow \theta_{i,k} - \eta_{i,k}\left[\lambda_{\mathbf{t}}\theta_{i,k} + \sum_{u\in U}\mathbb{1}_{u,i}^{R}\delta_u g'(\gamma_{u,i})\widehat{r}_{u,i}\theta_{u,k}\right]$$
$$+ \eta_{i,k}\left(\sum_{u\in U}\mathbb{1}_{u,i}^{R}\delta_u g'(\gamma_{u,i})r_{u,i}\theta_{u,k}\right) \quad (28)$$

To cancel out the negative part, we need

$$\theta_{i,k} - \eta_{i,k}\left[\lambda_{\mathbf{t}}\theta_{i,k} + \sum_{u\in U}\mathbb{1}_{u,i}^{R}\delta_u g'(\gamma_{u,i})\widehat{r}_{u,i}\theta_{u,k}\right] = 0$$

The proper step size is then

$$\eta_{i,k} = \frac{\theta_{i,k}}{\lambda_{\mathbf{t}}\theta_{i,k} + \sum_{u\in U}\mathbb{1}_{u,i}^{R}\delta_u g'(\gamma_{u,i})\widehat{r}_{u,i}\theta_{u,k}}$$

With this value of $\eta_{i,k}$, the negative part is cancelled and only the last term in Eq. (28) remains. Thus, we obtain the following multiplicative update rule.

Updating topic factor k of item i

$$\theta_{i,k} \leftarrow \theta_{i,k} \times \frac{\sum_{u\in U} \mathbb{1}^R_{u,i}\delta_u g'(\gamma_{u,i}) r_{u,i}\theta_{u,k}}{\lambda_{\mathbf{t}}\theta_{i,k} + \sum_{u\in U} \mathbb{1}^R_{u,i}\delta_u g'(\gamma_{u,i})\widehat{r}_{u,i}\theta_{u,k}} \tag{29}$$

We can proceed similarly to obtain the following update rules for the remaining user and item factors.

Updating brand factor b of item i

$$\beta_{i,b} \leftarrow \beta_{i,b} \times \frac{\sum_{u\in U} \mathbb{1}^R_{u,i}(1-\delta_u) g'(\gamma_{u,i}) r_{u,i}\beta_{u,b}}{\lambda_{\mathbf{b}}\beta_{i,b} + \sum_{u\in U} \mathbb{1}^R_{u,i}(1-\delta_u) g'(\gamma_{u,i})\widehat{r}_{u,i}\beta_{u,b}} \tag{30}$$

Updating topic factor k of user u

$$\theta_{u,k} \leftarrow \theta_{u,k} \times \frac{\delta_u \left[\sum_{i\in I} \mathbb{1}^R_{u,i} g'(\gamma_{u,i}) r_{u,i}\theta_{i,k} + \sum_{v\in U} \mathbb{1}^W_{u,v} g'(\omega_{u,v}) w_{u,v}\theta_{v,k}\right]}{\lambda_{\mathbf{t}}\theta_{u,k} + \delta_u \left[\sum_{i\in I} \mathbb{1}^R_{u,i} g'(\gamma_{u,i})\widehat{r}_{u,i}\theta_{i,k} + \sum_{v\in U} \mathbb{1}^W_{u,v} g'(\omega_{u,v})\widehat{w}_{u,v}\theta_{v,k}\right]} \tag{31}$$

Updating brand factor b of user u

$$\beta_{u,b} \leftarrow \beta_{u,b} \times \frac{(1-\delta_u)\left[\sum_{i\in I} \mathbb{1}^R_{u,i} g'(\gamma_{u,i}) r_{u,i}\beta_{i,b} + \sum_{v\in U} \mathbb{1}^W_{u,v} g'(\omega_{u,v}) w_{u,v}\beta_{v,b}\right]}{\lambda_{\mathbf{b}}\beta_{u,b} + (1-\delta_u)\left[\sum_{i\in I} \mathbb{1}^R_{u,i} g'(\gamma_{u,i})\widehat{r}_{u,i}\beta_{i,b} + \sum_{v\in U} \mathbb{1}^W_{u,v} g'(\omega_{u,v})\widehat{w}_{u,v}\beta_{v,b}\right]} \tag{32}$$

To update the topic dependency weights δ_u's, we still use Eq. (27). As long as δ_u's remain in [0, 1], all these update rules will guarantee the non-negativity of user and item factors. We thus ensure this by applying cut-off to bring any δ_u outside [0, 1] back to the range.

5 Experiments

In this section, we conduct experiments on three datasets **FL-IMDB**, **4SQDB** and **ACMDB**. The first goal is to evaluate the performance of SocBIT and SocBIT$^+$ against other state-of-the-art methods in the task of rating prediction. Secondly, we examine the learnt topic and brand factors from both SocBIT and SocBIT$^+$. Finally, we characterize the brand conscious users determined by the two models.

As later demonstrated in our experiments, the topics learnt by our models and other methods are similar. We thus place more emphasis on the evaluation of brand factors and brand-conscious users as these are novel contributions of our models.

5.1 Experimental setup

For rating prediction on the real-world datasets, we evaluate SocBIT and SocBIT$^+$ against RSTE, SoRec and LibFM models. For LibFM, the meta-information to be included are item brands. The evaluation metrics is RMSE. In all experiments, we set hyper-parameters $\lambda_{\mathbf{t}} = 0.001$, $\lambda_W = 1$, $\lambda_{\mathbf{b}} = 0.1$ and $\lambda_{\mathbf{d}} = 1$.

Given a dataset, we first used fivefold cross validation (CV) to determine an appropriate number K of topics for each model. For each user with at least 5 adoptions, we divided her adoptions evenly into fivefolds. We iteratively use each fold as a test set and the others as the training set. For those users with less than 5 adoptions, we put all of her adoptions into the training set. We then trained each model using different K's in the range [5, 15] and determined the best K based on the average RMSE over the different test folds. This value of K will be fixed across all the models in later experiments. We also performed manual analysis on topics learnt by SocBIT$^+$ and provided them in Sect. 5.5.

Secondly, we examined the models accuracy in *rating* prediction in Sect. 5.2. We looked at prediction accuracy of the models when they are applied on users with few observed adoptions. The purpose is to see how our models fare against others in *cold-start* scenario.

Thirdly, we evaluated the models' performance in *adoption* prediction using precision-at-N and recall-at-N metrics in Sect. 5.3. We vary N from 1 to the total number of items in test set to check that recall values actually converge to 1. However, we only present the recall values for N's from 1 to 100, which is practical for recommendation tasks.

Finally, we analyzed brand-conscious users learnt by our models for both datasets in Sect. 5.4. We validated the results on brand-conscious users by showing that the brands they adopt either have high price (for **4SQDB**), high h-Index (for **ACMDB**) and STARmeter ranking (for **FL-IMDB**).

5.2 Evaluation on rating prediction task

First, we determine the appropriate number K of topics for each dataset by varying K from 5 to 15 and performed fivefold Cross Validation (CV). The results are shown in Fig. 4a, b, c for **FL-IMDB**, **4SQDB** and **ACMDB** respectively. From the figures, we can see that the best value of K for both **ACMDB** and **FL-IMDB** is 10 and for **4SQDB** is 9. From the algorithm perspective, these K values are the best since larger values do not offer any significant reduction in RMSE on training set and even increase RMSE on test set (overfitting). We may also explain this based on characteristics of each dataset. For **4SQDB**, $K = 9$ offers best performance on test set because it matches the actual categorization of cuisines in Singapore into nine popular types, namely {American, Chinese, Indian, Italian, Japanese, Thai, Seafood, Breakfast, BBQ}. For **FL-IMDB** and **ACMDB**, we believe that $K = 10$ is the best value because it provides the categorization of items similar to the widely accepted categorization in reality. These values of K will be used across all the models in later investigations.

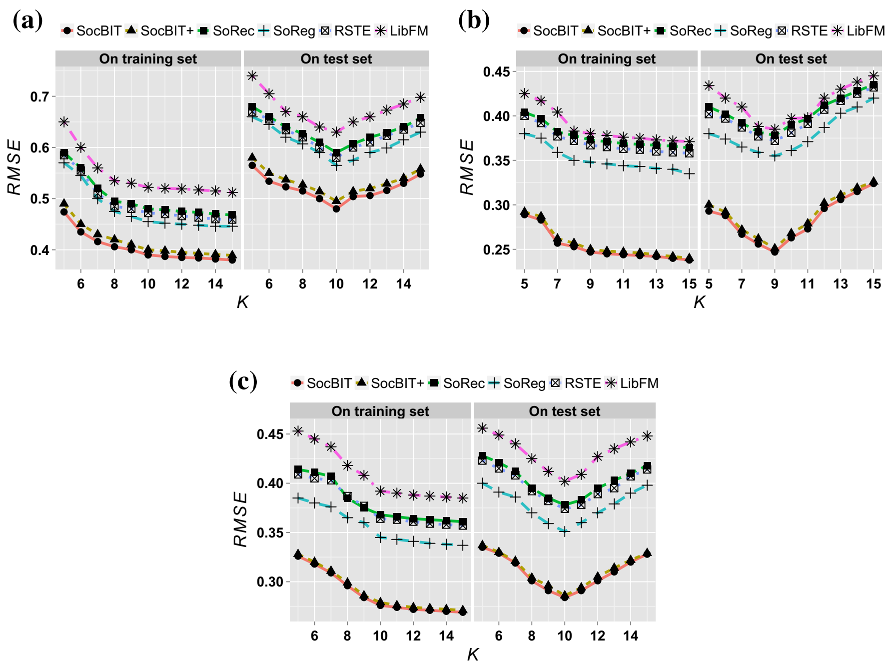


Fig. 4 Model RMSEs with respect to different K's ($p_{train} = 80\%$). **a FL-IMDB**. **b 4SQDB**. **c ACMDB**

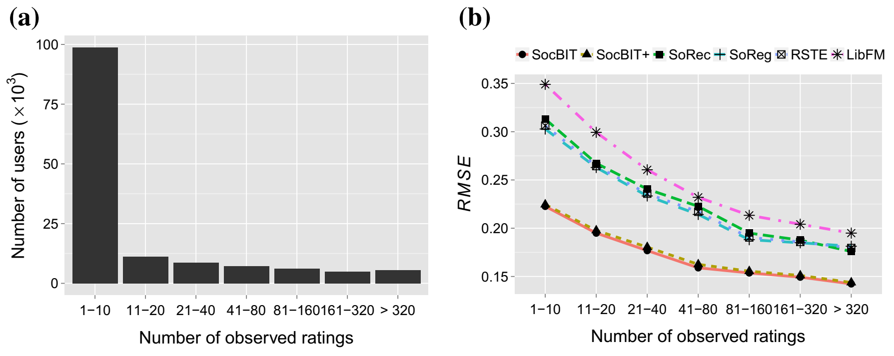


Fig. 5 **FL-IMDB**: accuracy of rating prediction w.r.t. number of observed ratings ($p_{train} = 80\%$). **a** Number of users in groups. **b** Test RMSEs for different user groups ($K = 10$)

5.2.1 Accuracy for different user groups

One of the challenges in recommendation systems research is to predict accurate ratings for a user even when she only rates a few items (i.e., *cold start* problem). We therefore want to investigate how well our model handles this challenge. For that purpose, we first group users based on the number of observed ratings in training data, and then evaluate prediction accuracies for different groups. We group users based on the distribution of rating count in each dataset (Figs. 5a, 6a, 7a) as follows.

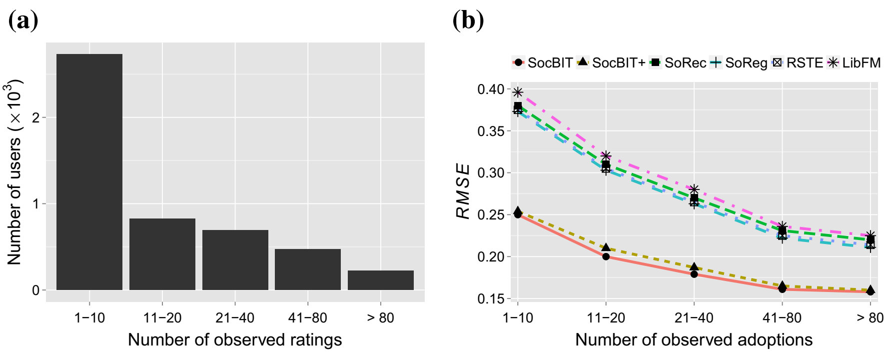


Fig. 6 **4SQDB**: accuracy of rating prediction w.r.t. number of observed ratings ($p_{train} = 80\%$). **a** Number of users in different groups. **b** Test RMSEs for different user groups ($K = 9$)

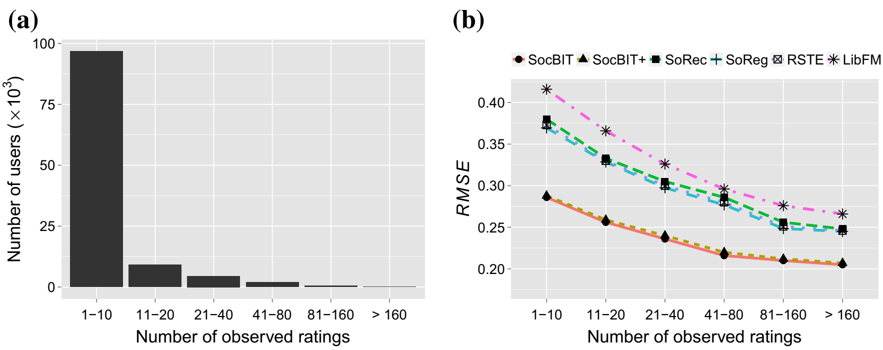


Fig. 7 **ACMDB**: accuracy of rating prediction w.r.t. number of observed ratings ($p_{train} = 80\%$). **a** Number of users in groups. **b** Test RMSEs for different user groups ($K = 10$)

– **FL-IMDB**: we form 7 groups, with rating counts in ranges [1, 10], [11, 20], [21, 40], [41, 80], [81, 160], [161, 320] and $(320, \infty)$.
– **4SQDB**: we form 5 groups, with rating counts in ranges [1, 10], [11, 20], [21, 40], [41, 80] and $(80, \infty)$.
– **ACMDB**: we form 6 groups, with rating counts in ranges [1, 10], [11, 20], [21, 40], [41, 80], [81, 160] and $(160, \infty)$.

From Figs. 5b, 6b, and 7b, we observe that SocBIT and SocBIT$^+$ perform equivalently well with SocBIT performs slightly better than SocBIT$^+$. This is expected as the space of parameters of the former covers that of the latter. Thus, the former should perform at least as good as the latter. More importantly, they both outperform other methods. Especially for users whose only few ratings can be observed, i.e. 1–10 ratings, our models offer much better prediction accuracy. The accuracy improvements of SocBIT$^+$ for this user group are as follows.

– On **FL-IMDB**: compared with SoReg, RSTE, SoRec and LibFM, SocBIT$^+$ reduces RMSE by 25.8, 26.7, 28.3 and 35.7% respectively.
– On **4SQDB**: compared with SoReg, RSTE, SoRec and LibFM, SocBIT$^+$ reduces RMSE by 31.8, 32.3, 33.1 and 35.8% respectively.

– On **ACMDB**: compared with SoReg, RSTE, SoRec and LibFM, SocBIT$^+$ reduces RMSE by 22, 22.8, 24.2, and 30.8% respectively.

5.3 Evaluation on adoption prediction task

We now evaluate our models' performance in predicting adoptions. For FL-IMDB data, we do not have explicit adoptions. We thus convert ratings to adoptions by considering that a user adopts an item when her rating for the item is larger than 0.5. This is similar to the method described in Cremonesi et al. (2008).

To measure precision and recall for each user u, we first use the trained models (with best K) to predict u's ratings for all items in test set. We then form a ranked list by ordering all the test items by their predicted ratings. We then pick top N items from the ranked list to form a top-N recommendation list for u. Finally, we compute the number of hits between this top-N list and the list A_u of test items adopted by u. The recall and precision at N for user u can then be computed as:

$$recall_u(N) = \frac{\#\, hits(N)}{|A_u|} \tag{33}$$

$$precision_u(N) = \frac{\#\, hits(N)}{N} \tag{34}$$

We take average over all users to get the average precision and recall at N, denoted as $precision(N)$ and $recall(N)$.

Figures 8, 9 and 10 report performance of models in adoption prediction task on **FL-IMDB**, **4SQDB** and **ACMDB** respectively. In terms of recall-at-N, Figs. 8a, 9a and 10a show that SocBIT and SocBIT$^+$ outperform other models significantly, especially for $N \geq 30$. In adddition, at $N = 10$ they achieve recall values about 0.2, i.e., SocBIT and SocBIT$^+$ have probabilities of retrieving 20% of movies adopted by a user using just top-10 movies recommended to her.

In terms of precision, SocBIT and SocBIT$^+$ also outperform other models. On **FL-IMDB** (see Fig. 8b), the two models achieve maximum precisions of 0.4 and 0.37 respectively while maximum precisions of SoRec, SoReg and RSTE are no more than

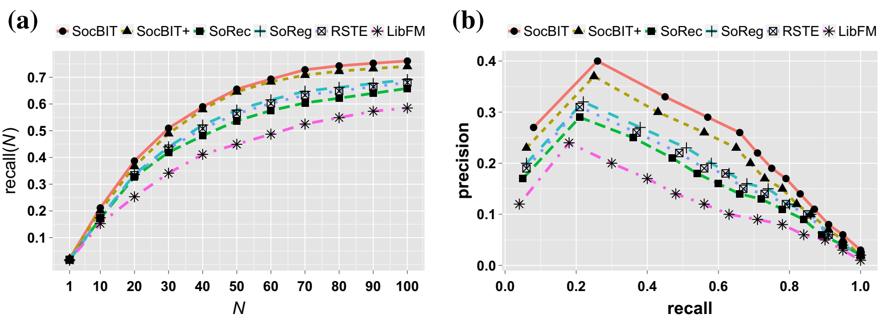


Fig. 8 **FL-IMDB**: models' performance in adoption prediction. **a** Recall-at-N. **b** Precision versus recall

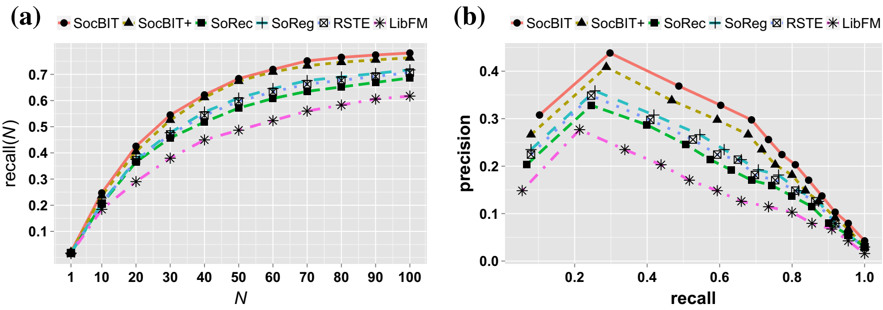


Fig. 9 **4SQDB**: models' performance in adoption prediction. **a** Recall-at-N. **b** Precision versus recall

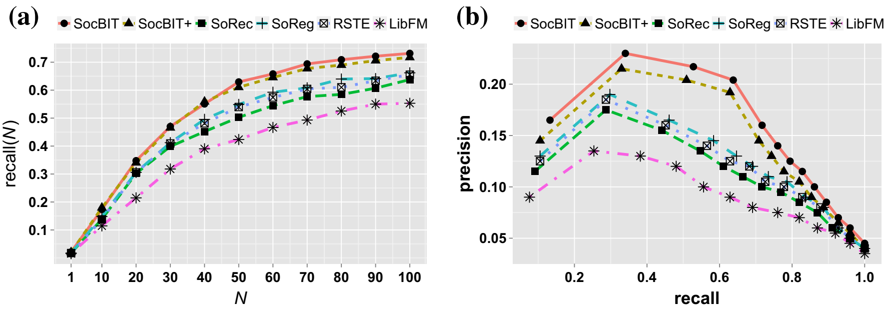


Fig. 10 **ACMDB**: models' performance in adoption prediction. **a** Recall-at-N. **b** Precision versus recall

0.32 and that of LibFM is just 0.24. On **4SQDB** (see Fig. 9b), the two models achieve maximum precisions of 0.44 and 0.41 while SoRec, SoReg and RSTE's maxima are no more than 0.36 and that of LibFM is just 0.28. Finally, on **ACMDB** (see Fig. 10b), the maximum precisions of SocBIT and SocBIT$^+$ are 0.23 and 0.21 while those of SoRec, SoReg and RSTE are no more than 0.19 and LibFM's precision can only reach 0.13.

5.4 Brand-conscious user identification

We now examine the brand-conscious users learned by SocBIT$^+$. Recall from Equation (11) that $1 - \delta_u$ represents inverse of topic dependency weight of user u. Thus, for each user u, $1 - \delta_u$ is a proxy for u's *brand-consciousness* level. We then look at the distribution of different brand-consciousness levels inferred by SocBIT$^+$ on the three datasets. As expected, the distributions follow a bell-curve form (see Figs. 11a, 12a, 13a) with about 80% users having medium brand-consciousness [in range (0.2, 0.8)] and only 20% of users having either low ($\leq$0.2) or high ($\geq$0.8) brand-consciousness.

To further confirm the validity of this brand-consciousness measure, we check its relationship with empirical measures. Precisely, we check if the more brand-conscious users are, the more likely they adopt items from established brands. For this purpose, we need an empirical measure allowing us to determine established brands. For **FL-**

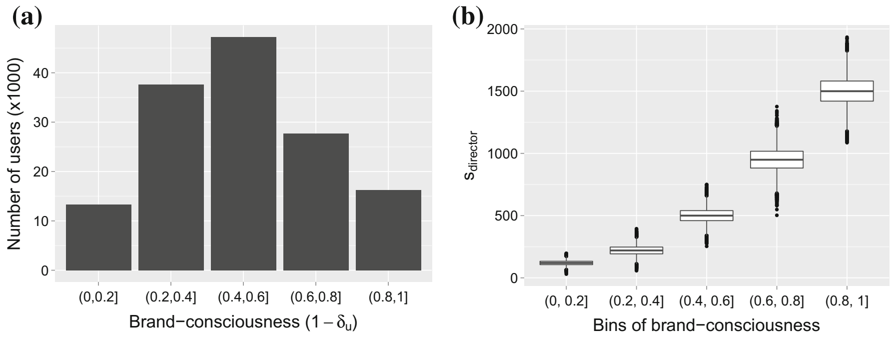


Fig. 11 Analysis of brand-conscious users identified in **FL-IMDB**. **a** Distribution of user brand-consciousness. **b** Score of directors adopted by users with different brand-consciousness

IMDB, we use *STARmeter rank of directors* (more details in Sect. 5.4.1). For **4SQDB**, we use *price*. For **ACMDB**, we use *h-Index*.

5.4.1 On FL-IMDB

We crawl the rank of directors given by *STARmeter*, which ranks directors and actors (note that the ranking is only available to users registering IMDBPro accounts). As IMDB restricts the maximum number of directors we can crawl, we first sample a subset of 2200 directors from the original set of ≈22 K directors. As this subset accounts for 10% of our brands, we believe that it is representative enough. We then crawl the ranks of directors in the subset. In the resultant set of directors, the highest rank is 27 (director Ben Affleck) and the lowest rank is 15,919 (director Roger Avary). We then plot the ranks against brand-consciousness level of users who watched movies of directors in this subset. Using ranks in plot however can cause confusion as the better a director, the smaller rank she has. We thus convert director ranks $r_{director}$ into director scores $s_{director}$ by a simple transformation $s_{director} = 16{,}000 - r_{director}$.

We divide users by their brand-consciousness into 5 bins, namely (0,0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8] and (0.8, 1]. For each bin, we plot score distribution of the directors adopted by users in the bin. The resultant plot in Fig. 11b shows that the more brand-conscious users are, the more selective they are in choosing movies to watch, they prefer to choose those from directors with high score.

5.4.2 On 4SQDB

We collect venue prices from a food review website *hungrygowhere.com* and obtain prices of about 80% of the number of venues in **4SQDB**. We then estimate the price of each brand as the average price of venues of the brand.

We divide users by their brand-consciousness into 5 bins, namely (0,0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8] and (0.8, 1]. For each bin, we plot its price distribution of the brands adopted by users in the bin. The resultant box plot given in Fig. 12b shows

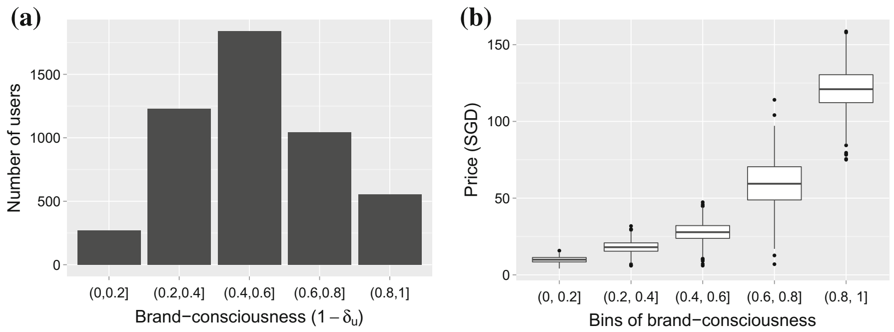


Fig. 12 Analysis of brand-conscious users identified in **4SQDB**. **a** Distribution of user brand-consciousness. **b** Prices of brands adopted by users with different brand-consciousness

Table 3 Top-5 brands adopted by brand-consciousness users versus those by normal users, the brands are sorted descendingly by adoption count

Top-5 brands adopted by users in BCU	Price	Top-5 brands adopted by normal users	Price
Punjab Grill	142	McDonald's	7
Pontini	142	Starbucks	10
Jaan	177	Swee Choon Tim Sum	13
Kaiseki Yoshiyuki	258	The Roti Prata House	6
Shinji by Kanesaka	335	Udders	6

All prices are in SGD

that the more brand-conscious users are, the more expensive brands they adopt. This matches our intuition.

Finally, we zoom in on the set of users who are highly brand-conscious, i.e., those with brand-consciousness more than or equal to 0.8. We denote the user set by BCU and compare them against normal users by the prices of the top-5 brands (in terms of adoption count) adopted by each user group. As shown in Table 3, users in BCU indeed adopt dining venues which are much more expensive than those adopted by normal user. The average price of brands adopted by the former is 163 (SGD) while that of brands adopted by normal users is just 8.4 (SGD). Moreover, all the top five brands adopted by users in BCU, are highly prestigious restaurants in luxury hotels or casinos in Singapore. These results match the intuition that highly brand-conscious users usually adopt expensive and/or prestigious brands. This again confirms that SocBIT^{+} can discover brand-conscious users in a reasonable manner.

5.4.3 On ACMDB

Similarly to previous experiment, we first divide users into 5 bins of brand-consciousness levels (0,0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8] and (0.8, 1]. We then crawl h-Index of brands, i.e., authors, from Google Scholar. To overcome the limit

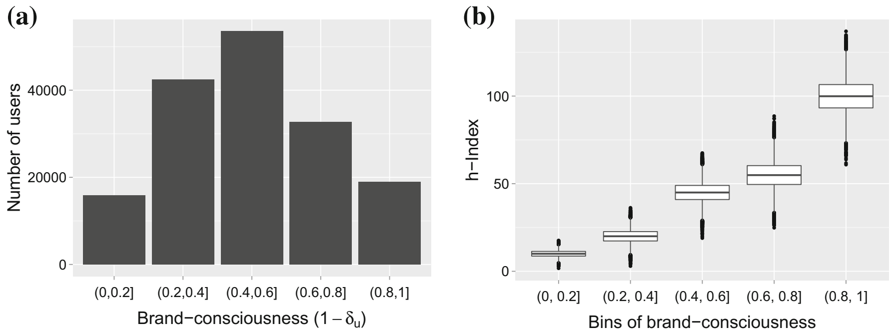


Fig. 13 Analysis of brand-conscious users identified in **ACMDB**. **a** Distribution of user brand-consciousness. **b** h-Indices of brands adopted by users with different brand-consciousness

Table 4 Learnt topics for **FL-IMDB** by SocBIT$^+$ and LibFM

Topic	Titles of top 5 movies SocBIT$^+$	LibFM
Action	The Dark Knight, Bourne Identity, Mr. and Mrs. Smith, Casino Royale, Quantum of Solace	The Bourne Ultimatum, The World is Not Enough, Tomorrow Never Dies, GoldenEye, Mr. and Mrs. Smith
Crime/Drama	The God Father (1, 2, 3), American Gangster, Gangs of New York, The Dark Knight, Taken	The Dark Knight, The God Father (2, 3), American Gangster, Double Indemnity, Casino
Sci-fi	Terminator 3, X-men, The Matrix(2, 3), Star Wars, Transformers: Revenge of the Fallen	I Robot, King Kong, X-men, Star Wars, Jurassic Park
Comedy	Dumb and Dumber, Borat, The Hangover, The 40-Year-Old Virgin, American Pie	American Pie, Meet the Parents, The 40-Year-Old Virgin, Borat, Harold and Kumar Go to White Castle
Romance	Titanic, Sleepless in Seatle, Roman Holiday, Gone with the Wind, You've got mail	Notting Hill, The English Patient, Roman Holiday, Breakfast at Tiffany's, Casablanca
Animation	Finding Nemo, Lion King, Wall-E, Toy Story, Monsters Inc	KungFu Panda, Mulan, Cars, 101 Dalmatians, Shrek
War	Saving Private Ryan, The Bridge on the River Kwai, Schindler's List, Apocalypse Now, The Thin Red Line	The Pianist, Enemy at the Gates, Apocalypse Now, Schindler's List, Pearl Harbor
Thriller	The Silence of the Lambs, The Sixth Sense, Taken, Primal Fear, Jaws	The Butterfly Effect, Jaws, Psycho, Basic Instinct, Seven
Fantasy	Lord of the Rings (1,2,3), The Chronicles of Narnia, Pirates of the Caribbean (1,2), Stardust	Lord of the Rings, Pirates of the Caribbean 1, Pan's Labyrinth, Spirited Away, Stardust
Family	Home Alone, Finding Nemo, Toy Story (1, 2), The Incredibles, Honey-I Shrunk the Kids	The Incredibles, Babe, Toy Story 1 Home Alone, Cars

The topics from LibFM are re-ordered to align with those from SocBIT$^+$

Table 5 Learnt topics for **4SQDB** by SocBIT^{+} and LibFM

Topic (cuisine)	Key words/phrases in top-10 venues	
	SocBIT^{+}	LibFM
American	Steak, fast food, hamburger, KFC, Starbucks, tavern, French fries	KFC, Mc. Donald, Astons Express, Swensen's, pancakes, drive-in
Chinese	Beijing roasted duck, chicken rice, yang chow fried rice, dim sum, mian (noodle)	Mian, yang chow fried rice, dim sum, si chuan food, porridge
Indian	Roti prata, Indian food, lamb curry, curry, punjabi chicken	Roti prata, curry, punjabi chicken, prawn curry, potato curry
Italian	Prego, Saizeriya, pasta, pizza, Oso Ristorante	PastaMania, pasta, Prego, pizza, Basilico
Japanese	Sushi, sashimi, ramen, udon, Pepper Lunch Express, Sakae sushi	Udon, sushi, bento, tempura, Ichiban sushi, My Izakaya, teriyaki
Thai	Thai Express, pineapple fried rice, Pad Thai, jasmine rice, green curry	Tom yum, basil rice, red curry, green curry, Pad Thai
Seafood	Seafood, chili crab, shrimp, fish-head steamboat, lobster	Pepper/chili crab, Korean seafood, fish-head steamboat, shark fin, shrimp
Breakfast	Pancakes, porridge, toast bread, half-boiled egg	Toast, omelet, coffee, fruit salad, bun
BBQ	BBQ, grill, Thai BBQ Korean BBQ, BBQ buffet	BBQ, Thai BBQ, BBQ buffet, Korean BBQ, outdoor BBQ

The topics from LibFM are re-ordered to align with those from SocBIT^{+}

on the number of authors we can crawl, we resort to using a sample for each bin. We randomly sample 200 authors adopted by users in the bin and collect the h-Index of those brands from Google Scholar. We then plot the distribution of those h-Indices. The results are shown in Fig. 13b. We observe that more brand-conscious users adopt higher h-Index authors. This is similar to our previous observations.

5.5 Topics learnt by SocBIT^{+} and LibFM

This section shows topics learnt by SocBIT^{+} and LibFM for the three datasets. On all the datasets, the two models agree on the set of topics and differ only by a permutation. For easy reading, we reorder the topics learnt by LibFM to align with those learnt by SocBIT^{+}. On **FL-IMDB**, we identify each topic by analyzing the top-5 movies under that topic. Each topic and the keywords in its top-5 movie names are provided in Table 4. The topics learnt for **FL-IMDB** turn out to be the following genres of movies: {Action, Crime/Drama, Sci-fi, Comedy, Romance, Animation, War, Thriller, Fantasy, Family}. Similarly, we identified each topic in **4SQDB** via the key words/phrases in top-10 venues under the topic (Table 5). The topics turn out to be the following cuisines {American, Chinese, Indian, Italian, Japanese, Thai, Seafood, Breakfast, BBQ}. Finally, we identified the ten topics in **ACMDB** based on title keywords of the top-10 papers of each topic. These topics and their keywords are provided

Table 6 Learnt topics for **ACMDB** by SocBIT$^+$ and LibFM

Topic	Key phrases in top-10 papers	
	SocBIT$^+$	LibFM
Database	Large DB, relational DB, key, joint operations, query, aggregation, (semi)-structured DB	Query, key, foreign key, DB, SQL, XML, (semi)-structured DB, structural join
Data mining	Mining, data, clustering, frequent patterns, k-means, classification, (un)supervised, SVM	Classification, regression, ensemble methods, random forest, boosting, association rule, k-means
Software engineering	Java, C++, development, path profiling, function/method calls, compiler, programming	C, C++, bug localization, dependency object-oriented, method, compiler, software
Internet	Network, lookup service, protocol, WWW, peer-to-peer, Internet, packet dynamics	WWW, Internet, network, latency, IP traceback, protocol, congestion, traffic
System	File system, performance, system design, caching, OS, (multi)processor, cache	Caching, system, architecture, cache, deadlock, processor, TinyOS, battery
Wireless/sensor network	Wireless, sensor networks, routing, directed diffusion, protocol	Router, routability, placement, distributed sensor networks, accurate
Distributed systems	Race detector, failure detectors, order, distribute systems, clocks, lock-free	Race, deadlock, workload, distributed consensus multithreaded, parallel, schedule
Security	Wireless security, cryptography, public-key, symmetric-key, digital signatures, anonymity	Privacy, protect, access control, anonymity, anonimous, public-key
Information retrieval	Index, inverted index, query, text/image retrieval, distributed IR	Inverted index, tf-idf, precision, recall, text retrieval, IR, query
Machine learning	Online learning, neural network, Bayes, image recognition, intelligent agents	Batch/statistical learning, Bayesian network, (un)supervised, machine translation

The topics from LibFM are re-ordered to align with those from SocBIT$^+$

in Table 6. The ten topics in **ACMDB** are the following ten research areas in computer science {database, data mining, software engineering, system, wireless/sensor network, distributed systems, security, IR, internet, machine learning}.

6 Conclusion and future work

In this work, we have demonstrated that mining brand-related factors, e.g. user *brand consciousness* and user *brand preference*, can provide actionable insights to recommendation tasks. The insights can be leveraged to make more accurate rating prediction. Moreover, we propose two novel probabilistic matrix factorization models, namely SocBIT and SocBIT$^+$, to incorporate such brand factors and social network information. Our experiments on real-world datasets show that SocBIT and SocBIT$^+$ achieve the following.

- Our models perform equivalently well and improve significantly accuracy of rating prediction over three state-of-the-art models: SoReg, RSTE and SoRec. Especially, for users with few observed ratings (1–10 ratings), both models offer much better accuracy in rating prediction. The improvements are at least 25.8, 31.8 and 22% on **FL-IMDB**, **4SQDB** and **ACMDB** respectively.
- Although SocBIT performs slightly better in rating prediction task, its gradient-based inference may return negative user and item factors, reducing interpretability. SocBIT$^+$ resolves this using a multiplicative inference, guaranteeing the non-negativity of learnt factors. The resultant factors learnt by SocBIT$^+$ are interpretable as topic and brand factors of users and items.
- SocBIT and SocBIT$^+$ also outperform other models in terms of precision- and recall-at-N. Especially for $N \geq 30$, the disparity is clear, e.g. recall-at-30 values of SocBIT and SocBIT$^+$ on **FL-IMDB**, **4SQDB** and **ACMDB** are at least 0.5, 0.52 and 0.46 respectively, recall-at-30 of other models are less than 0.45, 0.48 and 0.42 respectively. In terms of maximum precision, the improvements of our models over SoReg, RSTE and SoRec are 15.6, 13.8 and 10.5% on **FL-IMDB**, **4SQDB** and **ACMDB** respectively.
- SocBIT$^+$ infers user brand-consciousness from adoption data. The inferred brand consciousness scores follow a heavy tail distribution. Moreover, the more brand-conscious a user is, the more likely he adopts items from prestigious brands, characterized by expensive price on **4SQDB** or large h-index on **ACMDB**.

In this paper, we manually tune regularization coefficients λs. We thus plan to develop an automatic tuning method in future work. We also plan to combine user brand-consciousness and brand preference to infer whether a given brand is exclusive. Finally, when users are dependent on brand in item adoption, we can develop new methods to profile users based on information of the brands they adopt.

Acknowledgements This research is supported by the National Research Foundation, Prime Ministers Office, Singapore under its International Research Centres in Singapore Funding Initiative.

References

Ambler T (1997) Do brands benefit consumers? Int J Advert 16(3):167–198

Baltrunas L, Ludwig B, Ricci F (2011) Matrix factorization techniques for context aware recommendation. In: ACM conference on recommender systems (RecSys), pp 301–304

Bedi P, Kaur H, Marwaha S (2007) Trust based recommender system for semantic web. Int Joint Conf Artif Intell (IJCAI) 7:2677–2682

Belén del Río A, Vazquez R, Iglesias V (2001) The effects of brand associations on consumer response. J Consum Mark 18(5):410–425

Brewer MB (1991) The social self: on being the same and different at the same time. Personal Soc Psychol Bull 17(5):475–482

Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. In: ACM conference on knowledge discovery and data mining (SIGKDD), pp 160–168

Cremonesi P, Turrin R, Lentini E, Matteucci M (2008) An evaluation methodology for collaborative recommender systems. In: Automated solutions for cross media content and multi-channel distribution, pp 224–231
Ding CHQ, He X, Simon HD (2005) On the equivalence of nonnegative matrix factorization and spectral clustering. In: SIAM international conference on data mining (SDM), pp 606–610
Ding C, Li T, Peng W (2006a) Nonnegative matrix factorization and probabilistic latent semantic indexing: equivalence chi-square statistic, and a hybrid method. In: AAAI conference on artificial intelligence (AAAI)
Ding C, Li T, Peng W, Park H (2006b) Orthogonal nonnegative matrix t-factorizations for clustering. In: ACM conference on knowledge discovery and data mining (SIGKDD), pp 126–135
Erdem T, Keane MP (1996) Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. Mark Sci 15(1):1–20
Erdem T, Swait J, Broniarczyk S, Chakravarti D, Kapferer J-N, Keane M, Roberts J, Steenkamp JBEM, Zettelmeyer F (1999) Brand equity, consumer learning and choice. Mark Lett 10(3):301–318
Erdem T, Zhao Y, Valenzuela A (2004) Performance of store brands: a cross-country analysis of consumer store-brand preferences, perceptions, and risk. J Mark Res 41(1):86–100
Friedkin NE (2006) A structural theory of social influence, vol 13. Cambridge University Press, Cambridge
Graeff TR (1996) Using promotional messages to manage the effects of brand and self-image on brand evaluations. J Consum Mark 13(3):4–18
Hofmann T (2003) Collaborative filtering via Gaussian probabilistic latent semantic analysis. In: ACM conference on research and development in information retrieval (SIGIR)
Hofmann T (2004) Latent semantic models for collaborative filtering. ACM Trans Inf Syst (TOIS) 22(1):89–115
Hogg MK, Cox AJ, Keeling K (2000) The impact of self-monitoring on image congruence and product/brand evaluation. Eur J Mark 34(5/6):641–667
Hoyer PO (2002) Non-negative sparse coding. In: Neural networks for signal processing, pp 557–565
Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. J Mach Learn Res (JMLR) 5:1457–1469
Jamali M (2010) Flixster dataset. http://www.cs.ubc.ca/~jamalim/datasets/
Jamali M, Ester M (2010) A matrix factorization technique with trust propagation for recommendation in social networks. In: ACM Conference on recommender systems (RecSys), pp 135–142
Jiang M, Cui P, Liu R, Yang Q, Wang F, Zhu W, Yang S (2012) Social contextual recommendation. In: ACM conference on information and knowledge management, pp 45–54
Koren Y (2008) Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: ACM conference on knowledge discovery and data mining (SIGKDD), pp 426–434
Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. Computer 42(8):30–37
Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791
Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. In: Conference on neural information processing systems (NIPS)
Lin C-J (2007) Projected gradient methods for nonnegative matrix factorization. Neural Comput 19(10):2756–2779
Liu Z, Wang F, Zheng Q (2015) Modeling users' adoption behaviors with social selection and influence. In: Proceedings of the 2015 SIAM international conference on data mining, pp 253–261
Long MM, Schiffman LG (2000) Consumption values and relationships: segmenting the market for frequency programs. J Consum Mark 17(3):214–232
Ma H, Yang H, Lyu MR, King I (2008) Sorec: social recommendation using probabilistic matrix factorization. In: ACM conference on information and knowledge management (CIKM), pp 931–940
Ma H, King I, Lyu MR (2009a) Learning to recommend with social trust ensemble. In: ACM conference on research and development in information retrieval (SIGIR), pp 203–210
Ma H, Lyu MR, King I (2009b) Learning to recommend with trust and distrust relationships. In: ACM conference on recommender systems (RecSys), pp 189–196
Ma H, King I, Lyu MR (2011a) Learning to recommend with explicit and implicit social relations. ACM Trans Intell Syst Technol (TIST) 2(3):29
Ma H, Zhou D, Liu C, Lyu MR, King I (2011b) Recommender systems with social regularization. In: ACM conference on web search and data mining (WSDM), pp 287–296

Ma H, Zhou TC, Lyu MR, King I (2011c) Improving recommender systems by incorporating social contextual information. ACM Trans Inf Syst (TOIS) 29(2):9
Massa P, Avesani P (2007) Trust-aware recommender systems. In: ACM conference on recommender systems (RecSys), pp 17–24
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Soc 27:415–444
Mnih A, Salakhutdinov R (2007) Probabilistic matrix factorization. In: Conference on neural information processing systems (NIPS), pp 1257–1264
Paatero P, Tapper U (1994) Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2):111–126
Pilászy I, Zibriczky D, Tikk D (2010) Fast als-based matrix factorization for explicit and implicit feedback datasets. In: ACM conference on recommender systems (RecSys), pp 71–78
Rennie JDM, Srebro N (2005) Fast maximum margin matrix factorization for collaborative prediction. In: International conference on machine learning (ICML), pp 713–719
Salakhutdinov R, Mnih A (2008) Bayesian probabilistic matrix factorization using markov chain Monte Carlo. In: International conference on machine learning (ICML), pp 880–887
Shahnaz F, Berry MW, Pauca VP, Plemmons RJ (2006) Document clustering using nonnegative matrix factorization. Inf Process Manag 42(2):373–386
Solomon MR (1999) The value of status and the status of value. In: Holbrook MB (ed) Consumer value: a framework for analysis and research. Routledge, London, p 63–84
Srebro N, Rennie J, Jaakkola TS (2004) Maximum-margin matrix factorization. In: Conference on neural information processing systems (NIPS), pp 1329–1336
Su DL, Cui ZM, Wu J, Zhao PP (2013) Pre-filling collaborative filtering algorithm based on matrix factorization. Trans Tech Publ 411:2223–2228
Tang J, Gao H, Liu H (2012) mTrust: discerning multi-faceted trust in a connected world. In: ACM conference on web search and data mining (WSDM), pp 93–102
Tang J, Hu X, Gao H, Liu H (2013) Exploiting local and global social context for recommendation. In: International joint conference on artificial intelligence (IJCAI), pp 264–269
Tang J, Xia H, Liu H (2013) Social recommendation: a review. Soc Netw Anal Min 3(4):1113-1133
Ubilava D, Foster KA, Lusk JL, Nilsson T (2011) Differences in consumer preferences when facing branded versus non-branded choices. J Consum Behav 10(2):61–70
Vigneron F, Johnson LW (1999) A review and a conceptual framework of prestige-seeking consumer behavior. Acad Mark Sci Rev 1999:1
Wakita Y, Oku K, Huang H-H, Kawagoe K (2015) A fashion-brand recommender system using brand association rules and features. In: IIAI 4th international congress on advanced applied informatics (IIAI-AAI), pp 719–720
Weimer M, Karatzoglou A, Le QV, Smola A (2007) Maximum margin matrix factorization for collaborative ranking. In: Conference on neural information processing systems (NIPS), pp 1–8
Weimer M, Karatzoglou A, Bruch M (2009) Maximum margin matrix factorization for code recommendation. In: ACM conference on recommender systems (RecSys), pp 309–312
Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: ACM conference on research and development in information retrieval (SIGIR), pp 267–273
Yang CF, Ye M, Zhao J (2005) Document clustering based on nonnegative sparse matrix factorization. In: International conference on natural computation, pp 557–563
Zhang Y, Pennacchiotti M (2013) Recommending branded products from social media. In: ACM conference on recommender systems (RecSys), pp 77–84