

---

# Personalizing Information Using Users' Online Social Networks: A Case Study of CiteULike

Danielle Lee\*

---

## Abstract

This paper aims to assess the feasibility of a new and less-focused type of online sociability (the watching network) as a useful information source for personalized recommendations. In this paper, we recommend scientific articles of interests by using the shared interests between target users and their watching connections. Our recommendations are based on one typical social bookmarking system, CiteULike. The watching network-based recommendations, which use a much smaller size of user data, produces suggestions that are as good as the conventional Collaborative Filtering technique. The results demonstrate that the watching network is a useful information source and a feasible foundation for information personalization. Furthermore, the watching network is substitutable for anonymous peers of the Collaborative Filtering recommendations. This study shows the expandability of social network-based recommendations to the new type of online social networks.

## Keywords

CiteULike, Homophily, Information Personalization, Online Social Networks, Social Network-based Recommendations

---

## 1. Introduction

Due to users' active contributions in this Web 2.0 era, contents available on the Web continue to grow very rapidly [1]. The overwhelming amount makes it hard for users to locate useful information at the right time. Personalized recommendation technologies have emerged as one solution to cope with the information glut problem, and they have proven to be effective in a number of real-life applications, such as Amazon.com, Netflix, and Last.fm. Collaborative Filtering (CF) recommendation technology has especially received a lot of attention from both the world of academia and industry because of its relative simplicity and powerful performance. However, the CF technology is completely reliant upon the tastes of unknown users to make suggestions. Even though target users are the recipients of the recommendations, the ways to generate their CF-based recommendations is a black box, and the recommenders don't enable target users to get involved in and control the process. This lack of user control can directly decrease user satisfaction [2] and can cause a lot of problems, as detailed in Section

---

\* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Manuscript received January 06, 2014; accepted March 12, 2014.

**Corresponding Author:** Danielle Lee ([leedanie@u.washington.edu](mailto:leedanie@u.washington.edu))

\* Computing and Software Systems Program, University of Washington, Bothell, 18115 Campus Way NE, Bothell, WA 98011 ([leedanie@u.washington.edu](mailto:leedanie@u.washington.edu))

2. Therefore, the current recommendation technologies require finding a way to give users controls of their own recommendations. One effective solution is to utilize users' existing online social networks, rather than asking them for additional inputs and activities about their preferences. Accordingly, in contrast to the conventional CF technology, which is solely tailored by anonymous peers' tastes, recommendations based on users' online social networks has become one important evolution of the information personalization technology. It is a new generation of social network recommendations (hereafter, SN recommendations). The idea of taking advantage of users' social networks in personalizing their recommendations evolved from the Internet trend of online social networks, which are growing explosively. This trend is too important to be neglected.

Compared to the era when computer users were in isolation, Web 2.0 users have found it easier to establish contact with people of interest through online social networking sites (SNSs) such as Facebook, LinkedIn, MySpace, Orkut, etc. Moreover, beyond SNS, for which the main purpose is to network people, a variety of Web-based information management systems have adopted online social networks as a mechanism for users to acquire and share useful information. Naturally, users' self-defined online social links establish rich sources to disseminate various kinds of information (i.e., movies to watch, books to read, academic papers to refer to, bookmarks to explore, music concert to enjoy, etc.). The opinions and information gathered from the links influence social media users to a large extent [3-5]. According to preeminent social science theories—homophily and social influence—people selectively establish social links with those who are similar to them, and their attitudes, beliefs and behavioral propensities are affected by their social ties [6-8]. Therefore, with the proliferation of Web-based sociability, shared interests, and the resultant social influence, it is legitimate to expect for online social links to be merged with recommendation technology.

In spite of the abundant and diversified types of online sociability, SN recommendation technology is still in the infant stage. In particular, the summary of the relevant literature introduced in Section 2 indicates that most of the recent SN recommendations have concentrated on limited kinds of online sociality, which are trust-based networks and friendships. The relevant data sources are also restricted to only a few kinds of applications such as Epinions, FilmTrust, and Flixster. Due to the trouble of collecting data, instead of users' self-defined explicit social networks, sometimes researchers would make up implicit social networks via machine learning approaches. In order to meet these critical needs, this paper aims to investigate a newly emerging and less focused type of online social network (the watching network) as a component of personalized recommendations. We substantiated our approach with the less explored data source of CiteULike. The main purpose of this paper is to test the viability of the watching network as a useful foundation for personalized recommendations and to examine the substitutability of the watching network for anonymous peers of the conventional CF recommendations. We assessed the quality produced by the watching network-based recommendations by comparing it with typical CF recommendations.

## 2. Related Works

### 2.1 Collaborative Filtering Technology and Its Problem

CF technology emerged as an attempt to automate the word-of-mouth in the age of the Internet. The power of this technology is based on a relatively simple idea, which is as follows: begin with a target

user's ratings, find his/her cohorts that have similar interests to the target user, and recommend items favored by these cohorts to him/her. CF technology has proved its worth in recommending taste-based items such as movies, jokes, music, etc., where the preference is hard to be appreciated by the contents. It became popular for its ability to recommend serendipitous and diverse suggestions too. In spite of the strengths and big success in industry, a question has arisen regarding the quality of CF recommendations [2]. CF technology has turned out not to be well protected against malicious users who try to harm the system or to make a profit by gaming corresponding recommender systems. For example, a malicious user who copies the entire profiles of target users can mislead recommender systems into thinking the malicious user is a perfect cohort and then recommending his/her products to target users [9-11]. A group of ad-hoc users are also able to reinforce their own ratings and shift the recommendation predictions in the direction(s) that they desire so [12]. Even for well-intended users, if their tastes are eccentric (so-called 'black sheep users'), it is not easy to find their cohorts due to there being little overlap in tastes with other users [2]. CF recommender systems also suffer from other problems like the data-sparsity problem, cold-start users, and computational overload. In the situations where there are too many items to be rated in comparison with the total number of users or where the items last for a short period of time (e.g., job openings and news articles), the system struggles to find enough cohorts who adequately co-rated the same items as the target users [2, 13, 14]. For the cold-start users who just joined a system and rated too few items to infer their tastes, it is also very difficult to generate CF recommendations [2, 13, 15]. Moreover, because CF technology requires to compare one target user's taste with the tastes of all other users, it is computationally very expensive [15]. We suggest that these problems with CF technology occur in part because CF recommendations are utterly tailored by the tastes of anonymous peers who are selected by fully automated similarity calculations. Although users are the recipients of CF recommendations, systems don't allow them to get involved in the recommendation process. Even though the source of the recommendations is important criterion for judging the quality of recommendations, target users don't know who the peers are. Accordingly, they are unable to include or exclude their peers as they want [16]. As one way to get users to participate in their own recommendation process, we suggest that users' existing online social networks could be good sources of personalized recommendations.

## 2.2 Social Network-Based Recommendations

The idea to use social connections as a new source for personalized recommendations has emerged as a popular topic over the last a few years. Table 1 briefly summarizes the recent publications that focus on social network-based recommendations. This table only includes recommendations that utilize users' explicit social networks (i.e., users' self-defined networks) as a source of information. As the summary succinctly indicates, despite of the explosive popularity, abundance, and diversity of online sociability, the kinds of online social networks (trust-based networks and online friendships) used for personalized recommendations are highly limited. In addition, the types of target items and the data sources are also limited. For instance, most of the studies about trust network-based recommendations used Epinions.com. Unilateral online connections and professional collaboration networks that were taken into account for recommending research articles are a few exceptions [17-19]. Therefore, the goal of this paper is to fill in the missing parts by offering an extensive exploration of recommendation approaches that are based on less-focused online social networks.

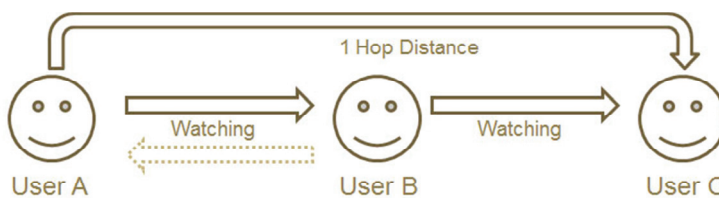
### 3. Watching Network: Social Links Centered on Objects

In this paper, this study examined the watching network as users' online social networks. A watching network is a typical example of newly emerging online sociabilities. Users in the Web 2.0 era have found it easier to see “who knows what” via their online social connections. It is a burden, however, to contact the person who knows the desired knowledge without any personal ties [20].

**Table 1.** Summary of Explicit Social Network-based Recommendation Studies

Paper	Kinds of social networks	Target items and the dataset
<b>Trust network-based recommendations</b>		
Avesani et al. [21]	FOAF and the trust values	Ski resorts (Moleskiing.it)
Al-Sharawneh & Williams [22]	Trust networks	General products (Epinions)
Chia & Pitsilis [23]	Trust networks	General products (Epinions)
DuBois et al. [24]	Trust networks	Movies (FilmTrust)
Golbeck & Hendler [25]	Trust networks	Movies (FilmTrust)
Guo et al. [26]	Trust networks	Movie (FilmTrust & Flixster) and General products (Epinions)
Jamali & Ester [27]	Trust networks	General products (Epinions)
Jamali & Ester [28]	Trust networks	General products (Epinions) & Movies (Flixster)
Ma et al. [29]	Trust networks	General products (Epinions)
Matsuo & Yamamoto [30]	Trust networks	Cosmetics (@cosme)
Mgureanu et al. [31]	Trust networks	General Products (Yahoo! Webscope & Epinions)
Massa & Avesani [10]	Trust networks	General products (Epinions)
O'Doherty et al. [32]	Trust networks & implicitly inferred trust-based links	Movie (FilmTrust & Flixster), Dating partners (LibimSeti), General products (Epinions)
Walter et al. [33]	Trust networks	General products (Epinions)
<b>Friends-based recommendations</b>		
Bonhard et al. [34]	Friendships	Movies (MovieMatch)
Bourke et al. [35]	Friendships	Movies/TV (Facebook)
Pera & Ng [36]	Friendships	Book (Amazon and LibraryThing)
Sinha & Swearingen [37]	Friendship	Books (Amazon, Sleeper & RatingZones) and Movies (Amazon & Reel.com)
Konstas et al. [38]	Friendship	Music (Last.fm)
Groh & Ehmgig [39]	Friendships	Local clubs (A German Site)
Liu & Lee [40]	Friendships	Online products (Cyworld)
<b>Other social networks-based recommendations</b>		
Lee & Brusilovsky [18]	Watching network (i.e. Unilateral Network)	Academic articles (CiteULike)
Lee & Brusilovsky [19]	Group membership	Academic articles (CiteULike)
Guy et al. [17]	Professional colleagues	Bookmarks of Internet/Intranet pages (Lotus Connections)
Yuan et al. [41]	Group memberships & friendship	Music (Last.fm)

Unlike online social networking systems that mainly link *mutually agreed-up* friendships, many social bookmarking systems, which aim to manage and share interesting information, offer users a new kind of social network without placing any burden on them of asking for mutual consent to be connected. They are allowed to ‘watch’ other users’ bookmark collections continuously, once users find the other users whose information collections are interesting. The watching relations don’t require any personal interactions and emotional bonds to make connections or even mutual agreements for being connected as watching connections. A *unilateral* relation is what forms a special kind of online social network. For instance, as shown in Fig. 1, User A is watching User B, but not necessarily vice versa. Hence, User B is a *watched party* of User A. The examples that are similar to the watching network are ‘following’ on *Twitter* and *Google Plus*, ‘network’ on *Delicious*, ‘contacts’ on *Flickr*, or ‘subscriptions’ on *Youtube*. This relation is convenient and purely based on the utility of information possessed by users being watched. In addition, it can compensate for privacy concerns because the watched users can decide what to expose or hide from strangers by themselves.



**Fig. 1.** Social links with distance.

Unilateral relations gained attention along with the success of social bookmarking and micro-blogging applications [42-44]. In the early part of this century, with the evolution of Web technologies, Wellman [45] suggested that various new online relationships would emerge, and that the networks would be ‘less bounded’. The unilateral relationship is one kind of the new types of online relationships and is exemplified by the watching network. Unlike reciprocal friendships on SNS (e.g., Facebook, MySpace, or Friendster) where users increase the number of friends simply for fun or curiosity [45], the watching relationships are designed on acquiring useful information from the watched parties’ collections. When a user unilaterally watches many users, it may cause rapid growth and potential dilution of the watcher’s list of information items collected from his/her watched parties. Therefore, unilateral relationships require users to be careful in selecting people to follow, based on the utility of information or taste.

Some researchers may argue that watching relations are not social networks, since they may be not based on any social interactions and emotional bonds. However, our preliminary works [46-48] found an important social property (homophily) in the relationships. Specifically, the relations meet the similarity attraction hypothesis [49] and held the transitivity power [50]. A high degree of similarity was embedded in unilateral relations and the similarity decreased with the increase of distance. Even though two users in a watching connection tend not to interact personally and not to share any emotions, they are directed networks located in the center of interesting information objects and naturally have highly object-centered sociality. Breslin and Decker [52] said that the social networks being connected via items of interests are more long-lived relations than the relationships that do not share any item of interest. In order to help users develop better relationships, the authors insisted that SNSs have to take

people's actions about contents (e.g., such as tagging, blogging, adding comments, etc.) into account to find out users' items of interests [51]. Even though many social bookmarking systems support these watching relations, there are surprisingly few studies to utilize this novel online network for personalizing users' information. Java et al. [53] insisted that one of the reasons why users enjoy micro-blogging, such as Twitter, is to share information. They also found three kinds of users' intentions for socializing on Twitter (i.e., *following*), which are not only 1) friendship-wise relationships, but also 2) information sharing and 3) information seeking.

We interpreted the watching connections existing on social bookmarking systems to mimic the process of bookmarking items of interest. In this context, users bookmark other users of interest. Therefore, the watched users' information can be used as a part of the watching users' preferences and furthermore, it can be used as a foundation for watching users to acquire useful information.

## 4. Recommendation Algorithms

Before examining recommendation algorithms, we need to consider hypotheses to be tested. Since the main focus of this paper is on the benefit of the watching network as a viable alternative to the anonymous peers, the watching network-based recommendations will be compared with the typical CF recommendations. Therefore, the most critical hypothesis is as explained below.

**H1. Watching network-based recommendations produce better suggestions than traditional CF approach.**

As mentioned above, the watching users declared their watching links based on the perceived utility of the watched parties' information collections. Then, since the perceived utility tends to be rooted in shared interests, the watched users could also take advantage of the information possessed by the people who are watching them. These characteristics of watching relationship bring up the following two questions. Is it better to include not only the connections to be watched by our target users but also the connections watching our target users? When target users have very few watching connections, instead of relying on the limited number of direct social links, is it effective to include social links connected with distances (for instance, the relations between user A and user C on the Fig. 1)? The hypotheses below examine these two questions.

**H2. Combining both direct watching and watched connections is a better recommendation source than directly watched connections.**

**H3. Combining direct and indirect watched connections is a better recommendation source than directly watched partners.**

Lastly, because the CF recommendations, which are based on the top N anonymous peers, are inclined to perform well, can hybrid recommendation fusing users' top N anonymous peers with their watching connections enhance the recommendation quality?

**H4. Social recommendations solely based on watching network are better than the hybrid recommendations fusing both watching network and anonymous top N peer users.**

Because the CiteULike is a typical bibliography management system, the final output of this recommendation is research articles in which various metadata, such as titles, author names, and journal or conference names, etc. is available. Moreover, in order to enable users to manage their favorite items more effectively and express their interests about the items in a cognitive way, the

CiteULike system offers functions for being able to annotate free-text tags while bookmarking items. As a way to augment the recommendation quality, it is critical to take advantage of the various metadata in the personalization process. Thus, in the design of the watching network-based recommendation algorithms, the following three aspects are considered: 1) users' bookmarks of research articles or books, 2) their watching network, and 3) the metadata of the bookmarked items. Fig. 2 depicts the overall design of recommendation approaches.

Throughout this paper, for the CiteULike dataset consisting of users' bookmarks,  $B$  is the user-item bookmark matrix of  $B=[B_{ui}]_{l \times n}$  where  $l$  and  $n$  denote the number of users and items, respectively.  $u \in \{u_1, \dots, u_l\}$  represents users and  $i \in \{i_1, \dots, i_n\}$  denotes items. The bookmark of user  $u$  on item  $i$  is  $b_{ui}$ . In here,  $\hat{b}_{ui}$  denotes a predicted bookmark of item  $i$  for user  $u$ , which is picked by our recommendation algorithm as a presumably favorable item.

		Recommendation Algorithms	
Selection of Peers	Anonymous Peers	• CF_CW	
	Social Peers	Direct Watching	• Watch_CW
		Direct Watching + Watched	• ReciWatch_CW
		Direct Watching + Indirect Watching	• 1hopWatch_CW
		Anonymous Peers + Direct Watching & Watched	• HybridSVD_CW

Fig. 2. Design of recommendation algorithms.

#### 4.1 Collaborative Filtering as a Baseline Recommendation Algorithm

The first recommendation algorithm is the conventional CF technology, as a baseline approach for this study. CF recommendations are based on item preferences of the whole user population in the CiteULike dataset. The CiteULike system doesn't provide a numeric rating mechanism for users to express their preferences about items. Instead, users can bookmark them. Hence, users' preferences can be encoded as unary ratings. Unary rating means the presence of bookmarks represents users' interests on corresponding items, but there is no degree of preference. The absence of bookmarks doesn't necessarily represent that users are uninterested in or dislike items, since there is a high probability for users not to discover items yet. For CF recommendation recommendations, using the unary ratings represented by users' bookmarks, we used a matrix factorization algorithm. Specifically, this study computed the singular vector decomposition (SVD) with Alternating-Least-Squares with Weighted-Regularization (ALS-WR) factorizer [52]. This SVD starts with a low rank approximation of the  $l \times n$  bookmark matrix  $B$ . When  $f$  is the number of latent factors to be extracted, we are able to derive the following two  $f$ -dimensional lower rank matrices:  $R \approx PQ^T$ . One matrix  $P$  presents the users' latent factor matrix, where  $P \in \mathbb{B}^{l \times f}$ . The  $u$ -th row of  $P$ ,  $p_u$  is the user  $u$ 's vector and represents how much user



$u$  is associated with  $f$  latent factors. Another matrix  $Q$  presents items' latent factor matrix, where  $Q \in \mathbb{B}^{n \times f}$ . For the  $i$ -th row of  $Q$ ,  $q_i$  is the item  $i$ 's vector and represents how much an item  $i$  possesses  $f$  latent factors. By computing the inner product between the two vectors, we can approximate the probability of the missing bookmarks,  $\hat{b}_{ui}$  (i.e., how likely a certain user  $u$  is going to bookmark item  $i$ ) [53, 54] as:

$$\hat{b}_{ui} = p_u q_i^T \quad (1)$$

The major challenge of SVD is to minimize the error of prediction,  $e_{ui} = b_{ui} - \hat{b}_{ui}$ . When there are many missing elements in the bookmark matrix, this typical SVD algorithm cannot find  $P$  and  $Q$ . One way to solve this problem is to fill out the missing values, such as, with the average rating of a user or an item. However, when a system fills out the missing values in a wrong way, even if it is possible to find the  $P$  and  $Q$ , the results are highly prone to overfitting [55]. As the solution for these data sparsity and overfitting problems in SVD, Zhou et al. [56] proposed the ALS-WR, which is to learn the model by fitting the existing bookmark records. For the first stage, from known bookmark records, we need to model  $P$  and  $Q$  with a weighted- $\lambda$ -regularization. The optimal model is defined as:

$$(P^*, Q^*) = \min_{P, Q} \sum_{(u, i) \in B} e_{ui}^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \quad (2)$$

Where  $B$  contains all  $(u, i)$  pairs for which the  $b_{ui}$  is known (the training set).  $\lambda$  is a constant to control the degree of regularization and is usually decided by cross-validation. By learning from the existing bookmark history, this modeling aims to minimize the regularized squared error [55].

Now it is the stage to compute alternating least squares (ALSs). ALS aims to minimize noises and possible errors on the above regularized model. The basic idea is, when both  $P$  and  $Q$  are unknown, to fix one of the matrices and compute another matrix again by solving a least squares problem. This algorithm alternates the steps by fixing one of both matrices. When  $Q$  is fixed, so as to re-compute the  $P$  matrix, ALS computes a separate ridge regression for each user. It takes the latent vector ( $q_i$ ) of items bookmarked by the user  $u$  as an input variable and the new values of his/her bookmarks ( $b_{ui}$ ) as an output variable.

$$A_u = Q[u]^T Q[u] = \sum_{i \in B_u} q_i q_i^T \quad (3)$$

$$d_u = Q[u]^T b_u = \sum_{i \in B_u} b_{ui} q_i \quad (4)$$

where  $A_u$  is the covariance matrix of user  $u$ 's item latent factor vector (having the number of factor  $f$ ), which is the input, and  $d_u$  is the input-output covariance vector. Finally, it finds the optimal  $p_u$  by the ridge regression as:

$$p_u = (\lambda l_u E + A_u)^{-1} d_u \quad (5)$$

$E$  is the  $f$ -dimensional identity matrix:  $E \in \mathbb{B}^{f \times f}$ .  $l_u$  is the number of items bookmarked by user  $u$ .



Similarly, when  $P$  is fixed, we can re-compute the  $Q$  matrix using the following ( $n_i$  is the number of user who bookmarked item  $i$ ) [57].

$$q_i = (\lambda n_i E + A_i)^{-1} d_i \quad (6)$$

The SVD algorithm requires determining the optimal algorithmic settings through cross-validation evaluation. The preliminary cross-validation evaluation showed that the SVD having 50 factors and  $\lambda=0.15$  produced the best results. In Fig. 2, the CF recommendations are depicted as ‘CF.’

## 4.2 Watching Network-Based Recommendations and Hybrid Recommendations

The main focus of this study is to assess the substitutability of watching links (i.e., social peers) for anonymous peers in recommendations. In watching relationships exemplarily depicted in Fig. 1, User B who is watched by User A didn’t initiate the relations and doesn’t know the fact that User A is watching him/her. Due to shared interests between user A and B, however, User A could reversely be a useful information source to User B. In addition, although the users connected with distance, such as where User A and User C shared less similar interests than directly watching connections, their similarities were higher than the values of randomly paired users [58]. Therefore, target users’ indirect watching connections could be a valuable source, as well. In order to determine the optimal scope of the watching links as a foundation of the personalized recommendations, we propose three strategies in the selection of *social peers*: 1) direct watching connections only (‘Watch’ in Fig. 2); 2) direct watching and watched connections (‘ReciWatch’); and 3) direct and indirect watching connections (‘1hopWatch’). Particularly, since the similarity was decreased along with the increase of social distance, we only included users with 1 hop distance (such as the relationship between User A and User C in Fig. 1) as indirect watching connections.

In the step to compute the prediction probability of watching network-based recommendations, the same matrix factorization algorithm was used for the CF recommendations. However, in the execution of the matrix factorization algorithm for the watching network-based approach, we built a separate sub-matrix that was solely made up of the target users’ bookmarks (in the training sets) and the bookmarks of their corresponding social peers for every target user. Due to much smaller and less sparse bookmark matrices, systems could generate the recommendations more efficiently. The data sparsity of the whole CiteULike dataset in our consideration is 0.9998, but the average data sparsity of watching network-based sub-matrices is 0.771 ( $\sigma = 0.30$ ). In the conventional CF recommendations, the standard matrix factorization approach feeds the whole bookmark data into the computation one time. Therefore, the recommender systems require a huge memory capacity and often complain about the heap size error. Considering the whole dataset, the CF algorithm requires a large number of latent factors (e.g., >500 for Netflix dataset [56]), and the large number is directly connected to the computational cost, as well. On the other hand, the watching network-based approach takes advantage of a relatively smaller size of bookmarks and focuses on a narrow scope of user population. Therefore, recommender systems don’t need to compute a large number of factors. In addition, whenever a target user updates his/her bookmark history or changes his/her watching connections, the system is able to rebuild a small sub-

matrix only for the corresponding target user and re-calculates the matrix factorization right away. In a situation where online users' participations are being exponentially increased, the large-scale SVD has a serious scalability problem [59-61]. The preliminary cross-validation evaluation showed that in the case of all watching network-based recommendations, the SVD having 24 factors and  $\lambda=0.15$  produced the best results.

Lastly, the hybrid approach is to compute personalized recommendations by fusing target users' top N anonymous peers with their direct watching and watched connections (i.e., HybridSVD). In here, the number of top anonymous peers (the variable N) is determined as many as each target user's social peers are. In particular, among various hybridization strategies, the mixed hybrid strategy was used [62]. Put differently, two sources of information, where one source is the target users' top N anonymous peers and the other source is their watching network-based social peers, were mixed as one sub-matrix as a foundation of the recommendations. Once the foundation was built, then the same matrix factorization algorithm was computed to obtain the hybrid recommendations.

### 4.3 Content Similarity Weight

In the above CF recommendations, we applied a sophisticated matrix factorization algorithm. Nonetheless, the user preference inputs for the algorithm were simple unary ratings. The unary ratings merely account for which items interest target users, not the degree of interests or the reasons for the interests. Thus, it is hard to produce more elaborate and accurate suggestions. In addition, in CF recommendations, the prediction probability of candidate items (i.e., what the chances are that the target users would like the candidate items) is computed by aggregating and averaging out the similarities of the peers who bookmarked the candidate items. When candidate items belong to the same set of peers, the probability values are exactly the same, although they are different items and contain heterogeneous metadata. The same problem occurs in the watching network-based recommendations. To solve this problem and to enhance the recommendation accuracy, at the final selection of candidate items for the CF and watching network-based recommendations, we considered another source of information—the metadata of candidate items. In particular, we used the title and social tags of each item (i.e., an academic paper or a book). Unlike unary rating-based user preferences, descriptive metadata exhibits users' cognitive understandings and various viewpoints about items [63]. The content similarity weight aims to calculate how close the descriptive metadata of candidate items is to the target users' preferences, which are inferred by the same types of metadata of their bookmarked items, including their social tags.

In order to compute the content similarity weight, we used the vector space model [64]. Before computing the weight, we modeled user preferences in a vector format using the metadata information of users' bookmark history. For bookmarked items, which are mainly academic articles and books, the possible metadata is titles, abstracts, author names, journal/conference names, etc. However, in the CiteULike dataset, the abstracts and journal/conference names were missing for more than half of the items. Additionally, the author name information in CiteULike has a serious ambiguity issue. Instead of these metadata types, titles and social tags of the bookmarks were chosen to model user preferences. We built two separate metadata-based user profile vectors, which are the title-based user profile and social tag-based user profile. As Mori et al. [68] suggested, words in article titles are the contextual

representations of the articles and express the topics succinctly. Hence, the title-based user profile calculates how much a given user is interested in certain contents and contexts. In addition, social tags are about users' personal viewpoints about items. Thus, users' social tags can serve as rich evidence about user conceptual understanding or the categorization of items.

As title-based user profiles, all title keywords of a user's bookmarked items were aggregated into one bag of keywords. Particularly, since we used a 10-fold cross-validation evaluation strategy (for more detailed information, refer to Section 5.2), users' partial bookmarks used as a test set are excluded in the user profiles, and the title words were collected only for the remaining training set. Then, for making an effective comparison, we applied text-processing techniques to the bag of title keywords. The terms were case-normalized to the lower-case, and all stop words were removed. We also applied the Porter stemmer so as to reduce word variation to its stems or roots [65]. Finally, the processed bag of title keywords was transformed into a title keyword vector consisting of keywords and the Term Frequency/Inverse User Frequency (TF/IUF) values. In this paper, we computed each user as a document. Therefore, the document frequency variable in the original TF/IDF formula was replaced with the user frequency indicating how many users bookmarked papers where their titles contained the keywords. We also built the social tag-based use profiles using all of the social tags of each target user in the exact same way with the title-based user profiles.

As the next step, once a list of candidate items were selected from CF recommendations or social recommendations, the system compared how the title keywords and social tags of these candidate items were similar to the target users' profiles. In the same way that the system built user profiles, the keyword and tag-based vectors of each candidate item were also built. The title-based item vector is made up of all of the terms appearing in the title of a candidate item and the TF/IDF values. The social tag-based item vector consists of the annotated tags of the corresponding item and the TF/IDF values. For the tag-based vector we did not limit the scope of the social tags to the users' watching links and included all of the social tags annotated to the corresponding items, except for target users' own tags. Then, the similarities between the user profiles and each item vectors were computed using the cosine similarity. Using the resultant similarity, the system made a ranked list of candidate items [66]. The following equation denotes how the content similarity weights were applied in the recommendations:

$$CF\_CW_{u,i} = \frac{\sum_{v \in p_u} SVD\_Prob_{u,v}}{V} \times \left( \left( \frac{k_{ui} - \mu_K}{\sigma_K} \right) + \left( \frac{t_{ui} - \mu_T}{\sigma_T} \right) \right) \quad (7)$$

The above equation shows CF recommendations using the content similarity weight (CF\_CW) of a candidate item  $i$  for the target user  $u$ . For the candidate item, the system calculated the cosine similarity,  $k_{ui}$ , between user  $u$ 's title-based profile  $k_u$ , and title-based vector of item  $i$ ,  $k_i$ . It also calculated the cosine similarity of,  $t_{ui}$ , of user  $u$ 's tag-based profile  $t_u$  and tag-based vector of item  $i$ ,  $t_i$ . Before unifying these two cosine similarities as one content similarity weight and in order to normalize the values, the system applied the standard score using the mean  $\mu$  and standard deviation  $\sigma$  of the corresponding similarity measure, respectively. The unified content similarity value was multiplied with the final prediction probability value of the matrix factorization calculation for the item  $i$ . In here,  $v$  denotes one of the user  $u$ 's anonymous peers ( $p_u$ ) and has the candidate item  $i$  in his/her bookmarks.  $V$  is the number of anonymous peers who have the candidate item  $i$  in their bookmarks. In watching network-based

recommendations with content similarity weight,  $V$  and  $v$  denotes the numbers of social peers and each social peer. The preliminary cross-validation evaluation indicated that adding the content similarity weight to the recommendations is consistently and significantly helpful in increasing the accuracy and completeness of the recommendations. Hence, this study only considered the recommendations with a content similarity weight as the suffix ‘\_CW’ of each recommendation approach, as shown in Fig. 2.

## 5. Experimental Evaluation

### 5.1 Data Source

For the evaluation of our algorithms, this study used bookmarks and the tag records of CiteULike, which is one of the leading social bookmarking systems for managing and sharing bibliographic information, along with *Bibsonomy*, *Mendeley*, and *Diigo*. CiteULike data used in this study is made and distributed by the CiteULike administrators and was backed up on May 15, 2011. The dataset contains article IDs, all users, users' bookmarks, dates and times of the bookmarks, and the social tags at the time when this dataset was made. Since it doesn't contain information about the metadata of articles (i.e., titles, authors, publication journals, and abstracts) and watching relations, the information was collected separately. Table 2 shows the descriptive statistics. This dataset has 94,388 CiteULike users and 3,210,960 articles, consisting of about 3.9 million bookmarks. 66.3% of the users ( $n=62,568$ ) had five bookmarks at most and 89.9% of the articles ( $n=2,885,833$ ) were bookmarked only by one user. The bookmark sparsity of this CiteULike data set is 0.9999, which is similar with the sparsity of Netflix [67, p. 170]. The users who are in watching relations (whether they are watching or being watched) are 11,439, which equals 12.1% of the entire user population. They formed 44,847 watching relations in total. 3,223 users watch other users and are the target users of this study. Each target user watches about 14 users on average and 77.3% of them ( $n=2,492$ ) watch five connections at most. 9,897 users are watched by other users, and 4.53 users watch them on average.

**Table 2.** Descriptive statistics about the CiteULike Watching dataset

Watching Network	No. of users in watching relations	11,439
	No. of relations	44,847
Information collections	No. of distinct articles	3,210,960
	No. of users	94,388
	No. of bookmarks	3,869,993
	Avg. no. of bookmarks per user	41.0 ( $\sigma=407.5$ )
	Avg. no. of tags per user	143.21 ( $\sigma=2408.3$ )
	Avg. no. of bookmarkers per article	1.21 ( $\sigma=1.4$ )
	Avg. no. of tags per article	4.21 ( $\sigma=7.2$ )

### 5.2 The Formal Evaluation

For the evaluation of the proposed recommendation algorithms, 3,223 users who watch other user(s) and have at least one bookmarked item were selected as target users. Their bookmarks were randomly

split into 10 equal-sized subsets so as to execute 10 fold cross validation. For each iteration, one of the 10 subsets was used as a test set and the remaining 9 sets were used as a training set. When target users have less than 10 bookmarks, the set of their bookmarks was split into  $X$  sets ( $X < 10$ ), where each set consisted of one bookmark. Recommendations were assessed by how many recommended items are actually in the test set as hits. This process was repeated with a different test set 10 times, and the hit numbers were averaged out.

As the evaluation criteria, because the CiteULike dataset doesn't contain numeric ratings, the traditional evaluation methods for information retrieval (precision and recall) were used. The precision aims to measure how precise the recommendations are, and the recall measures how complete the recommendations are. In order to see the patterns of differences among our proposed approaches and the baseline, we measured precision and recall separately and did not compute the harmonic value F1. Both precision and recall will be computed for the top 10, 5, and 2 recommendations ( $N=10, 5$ , and 2). The recommendations are usually displayed in a ranked list. Users expected that items in a higher rank would be more important than the other items in lower ranks. Therefore, the position of correct recommendations is a critical evaluation criterion of recommendations.

$$precision@N = \frac{\text{No. of correct prediction}}{\text{N of top N set}} = \frac{\text{test} \cap \text{top N}}{N} \quad (8)$$

$$recall@N = \frac{\text{No. of correct prediction}}{\text{size of test set}} = \frac{\text{test} \cap \text{top N}}{\text{test}} \quad (9)$$

## 6. Results

### 6.1 Comparison of Various Recommendation Approaches

Before comparing the quality of our proposed watching network-based approach with the baseline CF recommendations, we evaluated the varying ways to include watching connections in personalized recommendations. In Section 4 above, we designed to assess the approach based only on direct watching connections ('Watch\_CW') against more extended scopes of watching connections—direct watching and watched connections ('ReciWatch\_CW'), and direct watching and 1hop watching connections ('1hopWatch\_CW') using hypothesis H2 and H3. Fig. 3 displays the results of both the precision and recall. According to the One-way ANOVA test, there were no significant differences among the three approaches at all three ranks ( $F=2.78, p=0.06$  for top 10 precision;  $F=3.95, p=0.02$  for top 5 precision;  $F=3.85, p=0.02$  for top 2 precision;  $F=0.31, p=0.73$  for top 10 recall;  $F=0.08, p=0.92$  for top 5 recall; and  $F=0.18, p=0.84$  for top 2 recall). Specifically, the addition of 1hop watching connections as social peers did not enhance the quality, and the average precision and recall were lower than the other two approaches, albeit insignificantly. The related hypothesis H3 is rejected. Moreover, 'ReciWatch\_CW' produced as good suggestions as 'Watch\_CW' did and failed to significantly outperform the "Watch\_CW". Hence, the hypothesis H2 is also rejected. However, 'ReciWatch\_CW' recommended precise and complete suggestions for a larger number of target users than 'Watch\_CW.' That is to say, the coverage of 'ReciWatch\_CW' was higher. For that reason, we decided that the 'ReciWatch\_CW' is the best way to include the watching connections among the three proposed

approaches.

Given that 'ReciWatch\_CW' was chosen as our final watching network-based approach, as the hypothesis H1 specifies, is the quality of the watching network-based recommendations better than the traditional CF recommendations? Fig. 3 also depicts the answers in terms of precision and recall. According to the t-test, the recommendations based on direct watching and watched connections ('ReciWatch\_CW') were as precise and complete as the CF-based ones are ( $t=0.82, p=0.41$  for top 10 precision;  $t=-0.20, p=0.84$  for top 5 precision;  $t=-0.74, p=0.45$  for top 2 precision;  $t=2.03, p=0.04$  for top 10 recall;  $t=0.91, p=0.37$  for top 5 recall; and  $t=0.83, p=0.41$  for top 2 recall). Due to the insignificant difference between two approaches, the relevant hypothesis H1 was rejected. Considering that the CF approach requires all of the bookmark data for the computation, the 'ReciWatch\_CW' generated equivalently good suggestions with much smaller sub-matrices only consisting of target users' and their social peers' bookmark records. We interpreted this result to mean that CiteULike users are connected with their watching partners via the utility of the partners' information and genuinely similar information preferences. Conclusively, the watching network is a useful information source and a feasible foundation for information personalization. Furthermore, the watching network is substitutable for anonymous peers of CF recommendations.

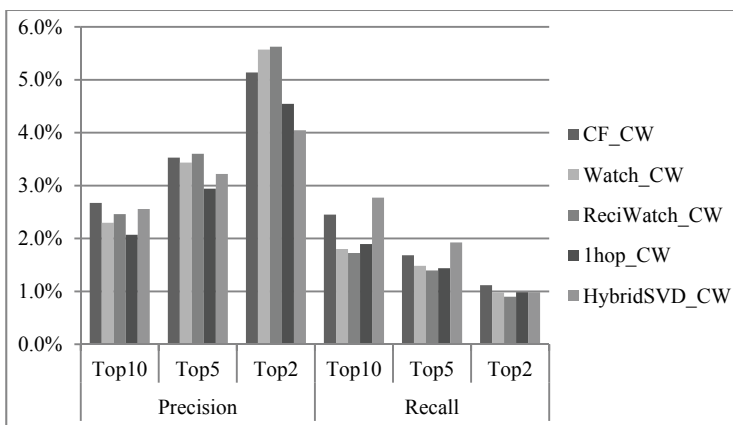


Fig. 3. Differences of recommendation quality depending on the kinds of recommendation approaches.

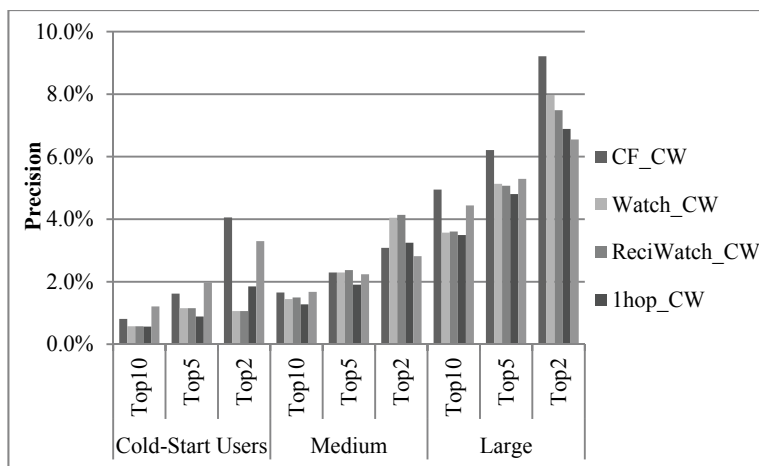
When the target users' anonymous peers and their watching connections are equally good information sources, can the combination of these two sources be better than either of them? To answer this question, taking the preferences of both the top N anonymous peers and watching connections into account generated hybrid recommendations. Then, the quality was compared with the CF and the watching network-based recommendations. The comparison of precision yielded no significant difference among the three approaches ( $F=0.39, p=0.68$  for top 10 precision;  $F=0.13, p=0.88$  for top 5 precision; and  $F=0.86, p=0.42$  for top 2 precision). The results of recall also showed insignificant differences ( $F=4.28, p=0.01$  for top 10 recall;  $F=1.50, p=0.22$  for top 5 recall; and  $F=1.09, p=0.39$  for top 2 recall). The relevant hypothesis H4 is rejected. That is to say, the combination of anonymous peers and social peers were not as effective as a source of information as we had expected. We interpreted this result to mean that, even though watching connection can substitute for anonymous peers, watching connections don't have any complementary abilities for anonymous peers.

## 6.2 Watching Network-Based Recommendation for Cold-Start Users

The descriptive statistics of the CiteULike data introduced in Section 5.1 showed that users participating in the watching network possess richer bookmark collections than the users who are not in a part of the network. However, about 10% of the CiteULike watching network participants had less than five bookmarks. Usually, in recommendations, users having this little amount of bookmarks or ratings are hard to receive reasonably good recommendations since the number of bookmarks is insufficient to properly present what they like. Therefore, these users are called cold-start users. One weakness of the CF recommendations is the cold-start user problem. Since the CF recommendations rely on the overlapped interests or tastes among users, when users have insufficient information about interests or tastes, it is difficult to guess what the presumably favorable items will be. As explained in Section 2, some researchers suggest that social network-based recommendations could be a good solution to solve this cold start user problem. It is because online social connections are rooted in the shared interests among users, as demonstrated in Section 6.1. Hence, although a target user doesn't have enough preference, the recommender system can include his/her online connections' preferences as a part of the target users' preference.

We will test whether watching a network-based recommendation is a good solution for a cold-start user problem or not. For this test, our target users were classified into three clusters according to the sizes of their bookmark collections—cold-start users ( $N < 5$ ), users having medium size bookmark collections ( $5 \leq N < 200$ ), and users having large collections ( $N \geq 200$ ). There were 316 users, 2,080 users, and 827 users in each cluster, respectively. For each user cluster, the best recommendation approach was examined.

Figs. 4 and 5 are the results of the precision and recall. In order to find out the statistical significance of the difference, we executed the two-way ANOVA test. The test demonstrated that the differences on the precision and recall were statistically significant according to the kinds of recommendation approaches and different user clusters ( $F=3.32, p<0.001$  for top 10 precision;  $F=2.83, p=0.002$  for top 5 precision;  $F=3.00, p<0.001$  for top 2 precision;  $F=5.55, p<0.001$  for top 10 recall;  $F=6.40, p<0.001$  for top 5 recall; and  $F=8.64, p<0.001$  for top 2 recall). As the next post-hoc analysis, we examined the patterns of difference among the various recommendation approaches for individual clusters of users.



**Fig. 4.** Differences of precision depending on users' number of bookmarks.



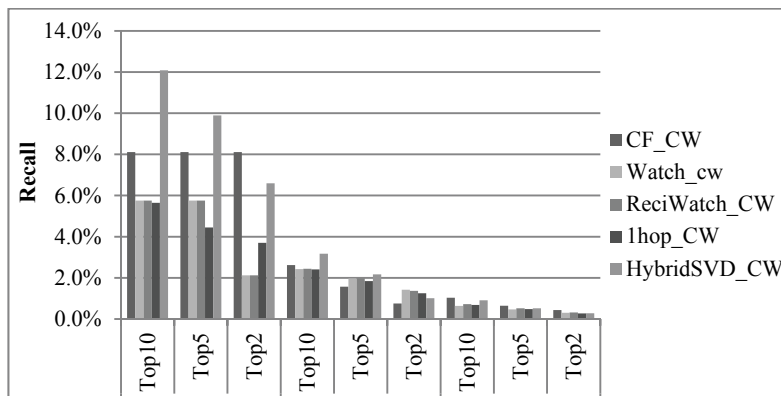


Fig. 5. Differences of precision depending on users' number of bookmarks.

First, for users that have a large bookmark collection, the differences between the various recommendation approaches are significant ( $F=9.02, p<0.001$  for top 10 precision;  $F=9.52, p<0.001$  for top 5 precision;  $F=11.37, p<0.001$  for top 2 precision;  $F=6.83, p<0.001$  for top 10 recall;  $F=7.21, p<0.001$  for top 5 recall; and  $F=8.74, p<0.001$  for top 2 recall). For this cluster of users, the CF recommendation is always the best option in terms of both precision and recall. Unfortunately, we found that none of the watching network-based recommendations generated good or better suggestions than the CF approach.

Second, for users that have a moderate number of bookmarks, the precision of various recommendation approaches were significantly different ( $F=5.11, p<0.001$  for top 10 precision;  $F=5.53, p<0.001$  for top 5 precision;  $F=7.73, p<0.001$  for top 2 precision), but the differences of recall were not significant ( $F=2.03, p=0.070$  for top 10 recall;  $F=1.06, p=0.378$  for top 5 recall;  $F=1.82, p=0.105$  for top 2 recall). Both the precision and recall indicated that the watching network-based approach was the best option for this cluster of users. In terms of precision, the watching network-based recommendations performed significantly better than the others, including the CF recommendations. Even though the differences were insignificant, the watching network-based recommendations also yielded the highest recall values on average.

Lastly, for cold-start users, the differences between the various recommendation approaches are significant for both precision and recall ( $F=4.06, p=0.001$  for top 10 precision;  $F=4.95, p<0.001$  for top 5 precision;  $F=5.19, p<0.001$  for top 2 precision;  $F=4.15, p<0.001$  for top 10 recall;  $F=5.00, p<0.001$  for top 5 recall; and  $F=5.83, p<0.001$  for top 2 recall). Specifically, the results, which are shown in Figs. 4 and 5, indicated that the hybrid approach is a good option for cold-start users. In lower ranks, the hybrid recommendations significantly outperformed the other approaches in both precision and recall. However, in the highest rank, the hybrid approach along with the CF recommendations performed the best without any statistical differences. Therefore, for cold-start users who have insufficient bookmarks, utilizing both anonymous users and users' social connections as the foundation for their recommendations is beneficial for increasing the quality of the information personalization.

## 7. Conclusions and Discussion

This paper examined the expandability of a social network-based recommendation to a new and less-

focused type of online sociability—the unilateral watching network. In this study, the empirical evaluation of various recommendation techniques focused on the CiteULike watching network and showed that our watching network-based algorithm was comparable in quality to conventional CF recommendations, regarding how the suggestions are accurate and complete. However, based on the fact that the watching network-based approach requires much smaller data, our proposed algorithm has better scalability than the CF recommendations. Viewed through the demonstrated results, CiteULike users built their watching connections through the shared interests among users. The watching network can provide a major benefit as a useful information source for the participants and a feasible foundation for information personalization. The result implies that the watching network can replace anonymous peers in personalized recommendations.

Lastly, according to the descriptive statistics, we found cold-start users whose bookmark collections were not enough to represent their preferences and that it is hard for them to receive reasonably good recommendations. For those cold-start users, the inclusion of users' social connections as one part of recommendations enhanced the quality, but it was not enough degree. It is more effective to take into account the watching connections as a complementary component of the CF recommendations as hybrid recommendations for cold-start users.

## References

- [1] J. Golbeck, "Introduction to computing with social trust," in *Computing with Social Trust*. London: Springer, 2009, pp. 1-5.
- [2] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web*. Heidelberg: Springer, 2007, pp. 291-324.
- [3] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, NV, 2008, pp. 7-15.
- [4] E. Bakshy, B. Karrer, and L. A. Adamic, "Social influence and the diffusion of user-created content," in *Proceedings of the 10th ACM Conference on Electronic Commerce*, Stanford, CA, 2009, pp. 325-334.
- [5] R. B. Cialdini and N. J. Goldstein, "Social influence: compliance and conformity," *Annual Review of Psychology*, vol. 55, pp. 591-621, 2004.
- [6] K. Carley, "A theory of group stability," *American Sociological Review*, vol. 56, no. 3, pp. 331-354, 1991.
- [7] S. R. Marks, "Intimacy in the public realm: the case of co-workers," *Social Forces*, vol. 72, no. 3, pp. 843-858, 1994.
- [8] J. C. Turner and K. J. Reynolds, "Self-categorization theory," in *The Handbook of Theories in Social Psychology (Volume 2)*. Los Angeles, CA: SAGE Publications, 2012, pp. 399-417.
- [9] S. K. Lam and J. Riedl, "Shilling recommender systems for fun and profit," in *Proceedings of the 13th International Conference on World Wide Web*, New York, NY, 2004 pp. 393-402.
- [10] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proceedings of the 2007 ACM Conference on Recommender Systems*, Minneapolis, MN, 2007, pp. 17-24.
- [11] B. Mehta, T. Hofmann, and W. Nejdl, "Robust collaborative filtering," in *Proceedings of the 2007 ACM Conference on Recommender Systems*, Minneapolis, MN, 2007 pp. 49-56.
- [12] J. O'Donovan and B. Smyth, "Trust in recommender systems," in *Proceedings of the 10th International Conference on Intelligent User Interfaces*, San Diego, CA, 2005, pp. 167-174.

- [13] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005.
- [14] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI-02)*, Edmonton, Canada, 2002, pp. 187-192.
- [15] P. Massa and P. Avesani, "Trust-aware collaborative filtering for recommender systems," in *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*. Heidelberg: Springer, 2004, pp. 492-508.
- [16] P. Bonhard and M. A. Sasse, "Knowing me, knowing you—using profiles and social networking to improve recommender systems," *BT Technology Journal*, vol. 24, no. 3, pp. 84-98, 2006.
- [17] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman, "Personalized recommendation of social software items based on social relations," in *Proceedings of the 3rd ACM Conference on Recommender Systems*, New York, NY, 2009, pp. 53-60.
- [18] D. H. Lee and P. Brusilovsky, "Improving recommendations using Watching Networks in a social tagging system," in *Proceedings of the 2011 iConference*, Seattle, WA, 2011, pp. 33-39.
- [19] D. H. Lee and P. Brusilovsky, "Using self-defined group activities for improving recommendations in collaborative tagging systems," in *Proceedings of the 4th ACM Conference on Recommender Systems*, Barcelona, Spain, 2010, pp. 221-224.
- [20] D. W. McDonald, "Recommending collaboration with social networks: a comparative evaluation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Lauderdale, FL, 2003, pp. 593-600.
- [21] P. Avesani, P. Massa, and R. Tiella, "A trust-enhanced recommender system application: Moleskiing," in *Proceedings of the 2005 ACM Symposium on Applied Computing*, Santa Fe, NM, 2005, pp. 1589-1593.
- [22] J. Al-Sharawneh and M. A. Williams, "Credibility-aware Web-based social network recommender: follow the leader," in *Proceedings of the 2nd Workshop on Recommender Systems and the Social Web (RSWeb)*, Barcelona, Spain, 2010, pp. 1-8.
- [23] P. H. Chia and G. Pitsilis, "Exploring the use of explicit trust links for filtering recommenders: a study on Epinions.com," *Information and Media Technologies*, vol. 6, no. 3, pp. 871-883, 2011.
- [24] T. DuBois, J. Golbeck, J. Kleint, and A. Srinivasan, "Improving recommendation accuracy by clustering social networks with trust," in *Proceedings of ACM Workshop on Recommender Systems & the Social Web (RSWeb)*, New York, NY, 2009, pp. 1-8.
- [25] J. Golbeck and J. Hendler, "Filmtrust: movie recommendations using trust in web-based social networks," in *Proceedings of the 3rd IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, 2006, pp. 282-286.
- [26] G. Guo, J. Zhang, and D. Thalmann, "A simple but effective method to incorporate trusted neighbors in recommender systems," in *User Modeling, Adaptation, and Personalization*. Heidelberg: Springer, 2012, pp. 114-125.
- [27] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, 2009, pp. 397-406.
- [28] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the 4th ACM Conference on Recommender Systems*, Barcelona, Spain, 2010, pp. 135-142.
- [29] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: social recommendation using probabilistic matrix

- factorization,” in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, Napa Valley, CA, 2008, pp. 931-940.
- [30] Y. Matsuo and H. Yamamoto, “Diffusion of recommendation through a trust network,” in *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*, Boulder, CO, 2007.
- [31] S. Magureanu, N. Dokoohaki, S. Mokarizadeh, and M. Matskin, “Design and analysis of a Gossip-based decentralized trust recommender system,” in *Proceedings of the 4th ACM RecSys Workshop on Recommender Systems and the Social Web*, Dublin, Ireland, 2012, pp. 1-8.
- [32] D. O’Doherty, S. Jouili, and P. Van Roy, “Trust-based recommendation: an empirical analysis,” in *Proceedings of the 6th ACM SIGKDD Workshop on Social Network Mining and Analysis (SNA-KDD)*, Beijing, China, 2012.
- [33] F. E. Walter, S. Battiston, and F. Schweitzer, “Personalised and dynamic trust in social networks,” in *Proceedings of the 3rd ACM Conference on Recommender Systems*, New York, NY, 2009, pp. 197-204.
- [34] P. Bonhard, M. A. Sasse, and C. Harries, “The devil you know knows best: how online recommendations can benefit from social networking,” in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it (BCS HCI 2007)*, University of Lancaster, UK, 2007, pp. 77-86.
- [35] S. Bourke, K. McCarthy, and B. Smyth, “Power to the people: exploring neighbourhood formations in social recommender system,” in *Proceedings of the 5th ACM Conference on Recommender Systems*, Chicago, IL, 2011, pp. 337-340.
- [36] M. S. Pera and Y. K. Ng, “With a little help from my friends: generating personalized book recommendations using data extracted from a social website,” in *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Lyon, France, 2011, pp. 96-99.
- [37] R. R. Sinha and K. Swearingen, “Comparing recommendations made by online systems and friends,” in *Proceedings of DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*, Dublin, Ireland, 2001.
- [38] I. Konstas, V. Stathopoulos, and J. M. Jose, “On social networks and collaborative recommendation,” in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, MA, 2009, pp. 195-202.
- [39] G. Groh and C. Ehmgig, “Recommendations in taste related domains: collaborative filtering vs. social filtering,” in *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, Sanibel Island, FL, 2007, pp. 127-136.
- [40] F. Liu and H. J. Lee, “Use of social network information to enhance collaborative filtering performance,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 4772-4778, 2010.
- [41] Q. Yuan, L. Chen, and S. Zhao, S. "Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation," in *Proceedings of the 5th ACM Conference on Recommender Systems*, Chicago, IL, 2011, pp. 245-252.
- [42] J. Brown, A. J. Broderick, and N. Lee, “Word of mouth communication within online communities: conceptualizing the online social network,” *Journal of Interactive Marketing*, vol. 21, no. 3, pp. 2-20, 2007.
- [43] M. Jandel and M. Elahi, “Tribal taste: mobile multiagent recommender system,” in *Proceedings of the 14th International Conference on Intelligent User Interfaces*, Sanibel Island, FL, 2009, pp. 489-490.
- [44] J. Preece and D. Maloney-Krichmar, “Online communities: focusing on sociability and usability,” in *The Human-Computer Interaction Handbook*. Mahwah, NJ: Lawrence Erlbaum Associate, 2003, pp. 596-620.

- [45] B. Wellman, "Computer networks as social networks," *Science*, vol. 293, no. 5537, pp. 2031-2034, 2001.
- [46] D. Lee, "How to measure the information similarity in unilateral relations: the case study of Delicious," in *Proceedings of the International Workshop on Modeling Social Media*, Toronto, Canada, 2010.
- [47] D. H. Lee and P. Brusilovsky, "Does trust influence information similarity?" in *Proceedings of the ACM RecSys Workshop on Recommender Systems & the Social Web*, New York, NY, 2009, pp. 71-74.
- [48] D. H. Lee and P. Brusilovsky, "Social networks and interest similarity: the case of CiteULike," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, Toronto, Canada, 2010, pp. 151-156.
- [49] P. R. Monge and N. S. Contractor, "Homophily, proximity, and social support theories," in *Theories of Communication Networks*. Oxford: Oxford University Press, 2003, pp. 223-239.
- [50] P. R. Monge and N. S. Contractor, "Network concepts, measures, and the multitheoretical, multilevel analytic framework," in *Theories of Communication Networks*. Oxford: Oxford University Press, 2003, pp. 29-77.
- [51] J. Breslin, "Social semantic information spaces," in *Semantic Digital Libraries*, S. R. Kruk and B. McDaniel, Eds., Heidelberg: Springer, 2009, pp. 55-68.
- [52] J. Breslin and S. Decker, "The future of social networks on the internet: the need for semantics," *IEEE Internet Computing*, vol. 11, no. 6, pp. 86-90, 2007.
- [53] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, CA, 2007, pp. 56-65.
- [54] X. Lin, J. E. Beaudoin, Y. Bui, and K. Desai, "Exploring characteristics of social classification," in *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*, Austin, TX, 2006, pp. 1-19.
- [55] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30-37, 2009.
- [56] X. Zhou, Y. Xu, Y. Li, A. Josang, and C. Cox, "The state-of-the-art in personalized recommender systems for social networking," *Artificial Intelligence Review*, vol. 37, no. 2, pp. 119-132, 2012.
- [57] D. Kalman, "A singularly valuable decomposition: the SVD of a matrix," *The College Mathematics Journal*, vol. 27, no. 1, pp. 2-23, 1996.
- [58] D. H. Lee and P. Brusilovsky, "Social networks and interest similarity: the case of CiteULike," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, Toronto, Canada, 2010, pp. 151-156.
- [59] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and Its Applications*, vol. 415, no. 1, pp. 20-30, 2006.
- [60] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Incremental singular value decomposition algorithms for highly scalable recommender systems," in *Proceedings of the 5th International Conference on Computer and Information Technology*, 2002.
- [61] S. Zhang, W. Wang, J. Ford, F. Makedon, and J. Pearlman, "Using singular value decomposition approximation for collaborative filtering," in *Proceedings of the 7th IEEE International Conference on E-Commerce Technology (CEC)*, Munich, Germany, 2005, pp. 257-264.
- [62] R. Burke, "Hybrid web recommender systems," in *The Adaptive Web*. Heidelberg: Springer, 2007, pp. 377-408.
- [63] C. Au Yeung and T. Iwata, "Capturing implicit user influence in online social sharing," in *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, Toronto, Canada, 2010, pp. 245-254.
- [64] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications*

*of the ACM*, vol. 18, no. 11, pp. 613-620, 1975.

- [65] B. Liu, "Informational retrieval and web search," in *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Heidelberg: Springer, 2007, pp. 183-236.
- [66] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: state of the art and trends," in *Recommender Systems Handbook*. New York: Springer, 2011, pp. 73-105.
- [67] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*. New York: Cambridge University Press, 2011.
- [68] J. Mori, T. Tsujishita, Y. Matsuo, and M. Ishizuka, "Extracting relations in social networks from the web using similarity between collective contexts," in *Proceedings of the 5th International Semantic Web Conference (ISWC)*, Athens, GA, 2006, pp. 487-500.



**Danielle Lee** <http://orcid.org/0000-0003-2106-716X>

Danielle Lee is an assistant professor at the Computing and Software Systems division, University of Washington Bothell. She received her Ph.D. degree in Information Science at the University of Pittsburgh and master's degree in Information Science at Syracuse University. She worked for Center for Dental Informatics, University of Pittsburgh as a researcher and Samsung SDS Co. as a software engineer. Her research interests are personalized recommendations using users' social networks, information propagation in online social networks and knowledge mining of collective intelligence.