Improving Personalized Recommendations using Community Membership Information

Danielle Hyunsook Lee

Adaptive Interactions, Co., suleehs@gmail.com, 1460-1 Bangbae-dong, Seocho-gu, Seoul, South Korea, 06568

Peter Brusilovsky
School of Information Sciences, University of Pittsburgh, peterb@pitt.edu, 135 N. Bellefield Ave., Pittsburgh, PA 15260

Abstract. While early recommender systems have mostly focused on numeric ratings to model their interests, recent research in this area has explored a range of other sources that can provide information about user interests, such as their bookmarks, tags, social links, or reviews. One source of information that has received little attention so far is users' membership in online communities. Online communities frequently evolve around specific topics. Therefore, user membership in a community could be interpreted as a sign of user interests in the topics of a particular community, and furthermore, could apply to personalized recommendations as a source of information. This paper explores the feasibility and the value of using users' community membership as a source of personalized recommendations for individual users. The first part of the paper focuses on feasibility. It attempts to assess to what extent the interests of users within the same community are truly similar. The second part focuses on the value of this information to personalized recommendations. It suggests several recommendation approaches that use community membership information. It also assesses the comparative quality of recommendations that are generated by these approaches. In particular, we substantiate our approach with one typical social bookmarking system, *CiteULike*. The results of our study demonstrate that the interests of members of the same communities are significantly closer than the interests of non-connected users. Moreover, we found that recommendation approaches based on community membership produce recommendations that are as accurate as those produced through a collaborative filtering approach, but with better efficiency. The recommendations are also more complete than those produced by a collaborative filtering approach. In addition, for cold-start users who have insufficient bookmarking information to reliably represent their interests, recommendations based on community membership are the most valuable.

Keywords: Social Network-based Recommendations, Personalized Recommendations, Online Community, Online Group, CiteULike

1. Introduction

The ability to create and join online communities has emerged as one of the most popular features in many types of social systems. Online communities usually form around recognizable topics, such as a fan club of a musician, a community of Hadoop programmers, an online forum for students taking the same class, or an online space for members of the same project. In this context, a user's membership in a community might indicate his or her interests in the topic of the community. Social dynamics in online communities extensively focus on contributing and distributing topic-relevant information [25]. Information shared by one community member frequently attracts the attention of other members [9]. Therefore, the social associations formed between members of the same community could be used as an information source to open up new possibilities to improve the information access of online users, and particularly to enhance personalized recommendations for users who are engaged in various communities.

To put this idea into the context of modern research about online sociality, we could consider membership in the same community as a social link between users. The pervasiveness of online sociality has brought scholarly attention to the use of online social networks as a valuable source of information for personalized recommendations. The direction of research is collectively referred to as 'social recommendations'. Social recommendations usually leverage users' online social networks by augmenting or replacing anonymous 'peers' used in collaborative filtering approaches with users' social connections. However, existing social recommendation approaches have predominantly[1] focused on just a few types of online social connections, such as friendship and trust [23, 24]. Despite the growing popularity of online communities (see Section 2.2), the social networks established by users' memberships in the same community have not been truly explored for generating social recommendations. This paper attempts to bridge this gap by examining the feasibility and value of community-based social networks as a useful information source for personalized recommendations.

The first part of this paper focuses on the feasibility of community membership as a useful information source. In specific, the parts attempts to uncover the presence of shared interests among the members of the same community. The presence of shared interests is a critical condition for using community membership information as a source of personalized recommendations. We examine the presence of shared interests by assessing the following hypothesis:

H.1 Information similarity between two members of the same community is higher than information similarity between two users who are not socially associated.

The second part focuses in the value of community membership information. The part investigates a range of approaches to generate personalized recommendations for individual users using their community membership. To assess the value of these approaches, we compare them to collaborative filtering (CF) by assessing the following hypothesis:

H.2 Recommendations based on users' self-defined community membership are better than collaborative filtering recommendations based on anonymous peers.

In this paper, both hypotheses are examined in the context of a popular social bookmarking system, *CiteULike,* where users actively participate in both communities and bookmarking activities.

The remainder of this article is structured as follows. Section 2 surveys existing literature on various problems of CF recommendation technology, online community membership, and other recommendation technologies based on online communities. Section 3 introduces the data set used in our study. The analysis of the shared interests among community co-members follows in Section 4. Section 5 introduces recommendation approaches based on community membership and assesses these approaches from several prospects. The article ends with a conclusion and discussion of possible areas of future work.

2. Related Work

2.1. Collaborative Filtering Recommendation Technology and Its Problems

Personalized recommendations have emerged as a solution to problems of information glut, which is caused by the overwhelming amount of information available on the web. Among various recommendation technologies (such as, content-based, case-based, demographic-based, hybrid recommendations, and so forth),

[1] According to our survey of the field, among the 40 existing studies of social recommendation approaches published through April 2016, 46% focus on friendship links and 39% focus on trust links between users [23].

the most popular is collaborative filtering (CF). A number of well-known companies, such as Amazon, Netflix, Last.fm, and YouTube have adopted and demonstrated the effectiveness of the technology [45, 48]. The CF systematically employs a process of 'word of mouth' to produce personalized suggestions based on preferences of like-minded anonymous 'peers' in a fully automated way. In spite of the strengths and big success of this approach, the use of a fully automated black-box process has called the quality of CF recommendations into question. There are several studies that have shown that CF technologies are vulnerable to attacks from malicious users [14]. For instance, a group of ad-hoc users are able to copy other users' rating profiles and shift the recommendation predictions to the desired directions to make profits [44]. Even if well-intended users have eccentric preferences (so-called "gray/black sheep users"), their small overlap of ratings with other users makes it difficult for a recommender system to find their peer cohorts and recommend relevant items [13]. CF recommendations also suffer from such problems as data sparsity, cold-start users, and computational overload [37]. The problems of CF technology have occurred in part because of a lack of user involvement in the recommendation processes. Although users are the recipients of CF recommendations, the recommender systems do not allow users to get involved in the choice of their peers—or even to know who their peers are. Therefore, it is timely to reconsider current CF technologies to find a way to allow users to participate in their recommendation process. In this paper, as one solution, we suggest substituting CF's anonymous "peers" with users' self-defined social connections, specifically community-based social connections.

2.2. *Online Community Membership*

Community activities are usually centered on one topic and are targeted to distribute and contribute topic-relevant information. Hence, we suggest that community membership embeds a strong object-centered sociality. The object-centered sociality theory explains that information objects are the main social interaction triggers and anchors of online communications [32]. Specifically, when a social system aims to manage information items, such as bibliographical information like *CiteULike*, the users' main motivations to join communities are to acquire and distribute useful information. For example, Hwang, et al. [15] demonstrated that the social dynamics of online communities are developed around information objects, rather than around social acquaintances or personal traits. The study showed that at the beginning of their activities, community members tend to rely more on the other members whose geographic and social distance is close to them. This behavior was caused by uncertainty about the given topic and concerns about sharing knowledge. However, as users accumulated more experience in the online community system and gained more knowledge, they preferred knowledge provided by experts [15].

The theory of *communal sharing* also indicates that the social dynamics of community membership are information-oriented. Community members treat information objects as their shared assets and are willing to contribute what they can. Members do not typically pay attention to the portion of contributions made by each individual member and do not expect to receive something in return for their contributions. Simply being a member of a community is sufficient, because they are able to use the resources that the community is sharing [1]. Unlike our study, which concentrates on group memberships that are explicitly defined by members, most of the studies about online community dynamics focus on deriving implicit communities using various machine-learning techniques and graph theories [3]. In this context, the contribution of our study is a novel personalized recommendation approach that is based on users' *explicitly defined* community memberships.

Beyond the context of personalized recommendations based on users' community memberships, a range of studies in different fields have demonstrated the importance of online communities as a source of useful information and knowledge. Some studies demonstrated how knowledge generated and shared within online communities brought positive outcomes to both companies (for example, in the development of new products, enhancement of customer satisfaction, and improvement of working environments) [38] and customers (such as increased brand perception, cost-efficient marketing, and improved customer loyalty) [36]. Researchers suggested that online communities could bring financial values to community providers [27]. In public health, online communities have emerged as critical sources for patients and caregivers to acquire and share health-related information [30]. However, existing research has rarely explored information-oriented interactions in online communities from individual members' point of views.

2.3. *Community-Based Recommendations*

The attempts to use community membership information in personalized recommendations could be classified into two considerably different streams. One stream focuses on generating recommendations for a community as a whole, rather than for individual members; while another stream exploits communities that are *implicitly* inferred by users' activities. Both streams are sometimes referred as community-based recommendation approaches.

Existing community-based recommendation approaches *for the whole community* suggest TV programs or movies to watch to a whole family [17]; venues and routes to travel to a group of people [46]; music to listen to when friends are together [21]; or restaurants that a group of friends would enjoy [2]. The recommendations targeting a group of people were typically generated by aggregating each member's preference or by integrating the suggested items for each member into a single set [4].

The idea of the second stream focused on *implicit communities* is to generate recommendations by first constructing implicit communities inferred by users' various information activities and then suggesting items of interest to individual users by using the implicit community information. Existing recommendations based on implicit communities suggest interesting learning contents to students [19], points of interest based on users' locations [42], or movies to enjoy [20].

Compared with a relatively large number of projects of the above two types, there are few projects that exploit users' self-defined community membership to generate personalized recommendations for individual members. The study by Yuan and colleagues is one of few examples where community membership was used as a foundation of recommendations for individual members. In this study, users' social connections, as defined by their community memberships, were fused with their online friendship connections. These integrated social "peers" were used as the foundations of social recommendations. A comparison with the CF approach using the Last.FM dataset, indicated that the use of social connections improved the overall quality of recommendations [43]. In another study, Garcia, et al. [12] generated recommendations of leisure activities at a tourist spot for groups of users, as well as for individual members of these groups using the same recommendation techniques. Moreover, the authors did not make any efforts to incorporate community membership information in the recommendations for individual users.

In the current study, we attempt to bridge the research gap indicated above by exploring the feasibility and the value of this *community-based sociality*. We performed an extensive two-part study of information encapsulated in users' self-defined community memberships as a source for personalized recommendations.

3. Data Source

As a context for our investigation, this paper examines *CiteULike.* CiteULike is one of the leading systems for managing and sharing bibliographic references. It is also one of the most well-known social bookmarking systems. Among other valuable features, the system enables users to engage in community activities. The system calls the communities 'groups'. As a result, throughout this paper, the terms 'community' and 'group' will be used interchangeably. When CiteULike users find interesting references to read, they are able to save the information in the form of 'bookmarks', both in their personal repositories and in their community spaces simultaneously. The community spaces are specifically labeled as "group library". Figure 1 briefly depicts the process. All other members of the corresponding community can see the newly posted bookmarks on their group library.

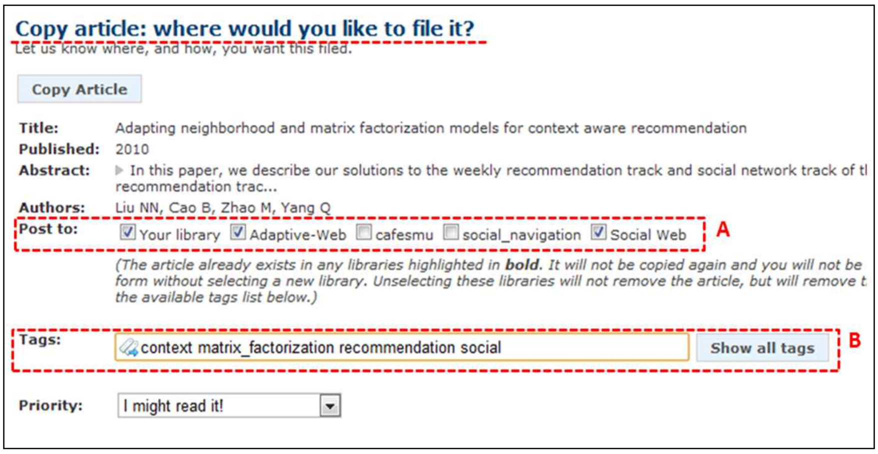

Figure 1. CiteULike interface enables users to post favorite references, both onto their personal repositories and group libraries simultaneously; the *A* area shows how a user can post an item to her personal repository ("your library") and her two groups ("Adaptive-Web" and "Social Web"). Social tags assigned to this item by this user (the *B* area) will be added to the item in both the personal repository and the group libraries.

The CiteULike data used in the study was assembled and distributed by the CiteULike administrators. The dataset contains the list of groups and members of each group. The dataset also includes a snapshot of each group library and each member's personal bookmark repository, as of the time when the dataset was made. Additionally, the dataset contains the other CiteULike users who did not participate in any community activities and their bookmark repositories. Both group libraries and individual users' repositories (whether they are group members or not) include the IDs of bookmarked references, social tags, and the timestamps. The metadata of the bookmarked references, such as titles, authors, published journals/conferences, abstracts, and publication years, was missing in the dataset. As a result, we collected the metadata separately. In the resulting dataset, in addition to the list of the whole CiteULike groups and the overall membership, both group libraries and individual users' personal repositories comprise the bookmarked references. Regarding each bookmarked reference, the resulting dataset also contains the timestamps and various types of metadata, which ranges from titles, authors, publication years, published journals/conferences, abstracts, and the social tags of the references. Table 1 presents these statistics in a descriptive form. The dataset included 94,388 users and 1870 groups. Note that 8.5% of the user population (i.e., 8009 members) is a member of groups. These 8,009 users who participated in the online communities are our target users.

Table 1. Descriptive Statistics about the Dataset of CiteULike Group Membership

Groups & the Memberships	No. of Groups	1,870
	No. of Users having at least one group membership	8,009
	No. of Group Membership	11,863
	Avg. No. of Members per Group	6.34 (σ = 16.3)
	Avg. No. of Memberships per Member	1.48 (σ = 1.7)
Information Collections	No. of Distinct Articles	3,210,960
	No. of Users	94,388
	No. of Bookmarks	3,869,993
	Total No. of Bookmarks in Group Library	324,486
	Total No. of Bookmarks in Members' Repositories	1,122,121
	Avg. No. of Bookmarks per Group	173.52 (σ = 784.1)
	Avg. No. of Bookmarks per Member	140.11 (σ = 1133.8)

4. Shared Information Preferences among Members of the Same Community

It is essential to determine if members of a same community (referred as co-members) share similar interests. The presence of shared interests among co-members is a required condition for the second part of this study—

community membership-based recommendations. This is required because the personalization is based on the overlapping preferences between a targeted user and their reference peers. To assess the presence of shared interests among community members, this section compares the information similarity of group co-members with other pairs of users who are not socially associated, as stated in H.1. Moreover, group libraries, which are jointly built by group members, might be good sources of information. Thus, we will assess the comparative advantage of group libraries over co-members' personal repositories as a practical information source.

In this section, first we describe the kinds of information similarity measures. Specifically, we consider two types of similarity measures: one type is based on users' bookmark records and another type is based on the metadata of users' bookmarks. The next sub-section presents the information-sharing pattern among community co-members. It is followed by an analysis to show the usefulness of both the personal repositories of co-members and of group libraries.

4.1. Information Similarity Measures

As explained above, the CiteULike dataset consists of bookmark collections and their related metadata. In order to include both types of data as users' information preferences, we explored seven kinds of bookmark-based and metadata-based similarity measures, as shown in Fig. 2. In the following sub-section, we explain how to compute each similarity measure.

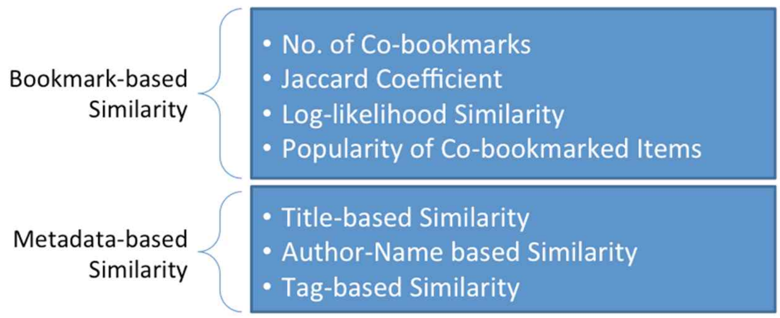

Figure 2. A summary of similarity measures

We will use the following notations for the rest of this paper. B is the user-item bookmark matrix, $B = \{B_{ui}\}_{l \times n}$ where l and n denote the number of users and items, respectively. $u \in \{u_1, \ldots, u_l\}$ represents users and $i \in \{i_1, \ldots, i_n\}$ denotes items. The bookmark of user u on item i is b_{ui}. The values of all bookmarks are unary (the presence of a bookmark only represents a user's interest).

4.1.1. Similarity Measures for Bookmarks

The first four kinds of measures gauge the degree of information similarity, based on users' bookmarks. The most traditional way to measure the similarity of two users' bookmark collections is *the number of co-bookmarks*. The measure simply counts the number of common items between two given bookmark collections. In spite of its popularity, this measure's major shortcoming is its absence of a tendency to proportionally compute the similarity to the size of two users' union bookmark set [26]. Let's assume that there are two user pairs. The number of union bookmark set made by one pair (e.g. user a and b) is 300 items, and the union bookmark set of another pair (e.g. user c and d) is 40 items. When the numbers of co-bookmarks of both pairs are 20, even though the similarities are the same, the former pair's agreement among bookmarks is much more significant than the latter pair.

To account for the varied sizes of union bookmark sets, the second bookmark-based similarity measure is the *Jaccard coefficient.* It is the most popular *normalized* measure for bookmarks [28, p. 401]. Let us assume that users *a* and *b* are in comparison and that variables B_a and B_b denote the users' bookmark collections, respectively. The Jaccard coefficient is computed as the following.

$$\text{Jaccard Power } (a, b) = \frac{|B_a \cap B_b|}{|B_a \cup B_b|} \qquad \text{Eq. (1)}$$

The third bookmark-based similarity is the *log-likelihood similarity*. This method considers how likely the overlap of interests are to occur by chance [29]. For example, co-bookmarking the same 10 items out of a collection of 100,000 items is much less likely (and thus could be considered as a stronger evidence of similarity) than co-bookmarking the same 10 items in a collection of 100. This measure also accounts for the number of items bookmarked by each of the given users and the number of items that are not bookmarked by either of them.

$$\Lambda(a, \mathrm{b}) = \frac{p_{a,b}}{p_b} \times \frac{p_{\sim a,\sim b}}{p_b} = \left(\frac{n_{a,b}}{n_a} \times \frac{n_b}{n_{a,\sim b}}\right) \times \left(\frac{n_{\sim a,\sim b}}{n_b} \times \frac{n_a}{n_{\sim a,b}}\right) = \frac{n_{a,b} n_{\sim a,\sim b}}{n_a n_b} \qquad \text{Eq. (2)}$$

where $\Lambda(\mathrm{a}, \mathrm{b})$ denotes the log-likelihood similarity between user a and user b. $n_{a,b}$ is the number of items bookmarked by both user *a* and *b*, and n_a or n_b is the number of items bookmarked by only user *a* or user *b*, respectively. $n_{\sim a,\sim b}$ is the number of items bookmarked by neither user. The resulting value is in the range 1 to -1, where a large positive value means that users are similar to one another, and a large negative value means that they are dissimilar to one another [29].

The last bookmark-based similarity measure is the *popularity weight* of co-bookmarked items. It tries to account for the *popularity bias* in measuring information similarity, given the fact that users are more likely to notice popular items than rare items [16]. Due to the popularity bias, co-bookmarking of popular items (for example, "Avatar" for measuring movie tastes) offers less evidence about users' shared interests than co-bookmarking of rare items (for example, an independent French film). Therefore, we incorporate popularity bias into the number of co-bookmarks by using the average inverse item popularity, as

$$\boldsymbol{P}_{ab} = \left(\sum_{i \in B_{ab}} \log\left(\frac{L}{l_i}\right)\right) / n_{ab} \qquad \text{Eq. (3)}$$

where IP_{ab} is the popularity weight of co-bookmarked items between user *a* and *b*. Item *i* is one of the items that users *a* and *b* have commonly bookmarked (i.e. $i \in B_{ab}$), and *L* is the total number of users. l_i is the number of users who bookmarked the item *i* and n_{ab} is the number of co-bookmarks between users *a* and *b*. Two users sharing less popular items are thus assigned a higher popularity weight value [39].

4.1.2. Similarity Measures for Metadata

Due to the overwhelming volume of information available online, our bookmarking process is irregular and opportunistic. It means that users with similar interests may not necessarily end up bookmarking exactly the same items. In this case, we can more reliably measure the similarity of interests on the level of an item's metadata than on the level of bookmarks. The three measures introduced in this subsection assess the degree of similarity between users using the metadata of users' bookmarked references. Given the nature of the CiteULike dataset, the kinds of metadata include reference titles, author names, and social tags. Note that our CiteULike dataset also contains other types of metadata, such as abstracts and journal/conference names, which could be used to infer users' preferences. However, CiteULike does not include these types of metadata for more than the half of the references, so we have left the exploration of them for future studies. In order to compute the similarity of these kinds of metadata, we used a vector space model [34].

The first metadata-based similarity is *title-based similarity*, which uses reference titles. As Mori et al. [31] suggested, a title of an article is a concise descriptor of the topic and title keywords identify semantic entities in articles. When title keywords in two articles are similar, the co-appearance of the same keywords represents the closeness of the semantic context among the entities of two articles [31]. When computing similarity based on titles, we collected the titles of all of the references in each user's bookmark repository. Then, all of the keywords of the titles were aggregated into one bag. Hence, we made 8,009 individual bags of title keywords for all of our target users. Then, for effective comparison, we applied text-processing techniques separately to each user's bag of title keywords. The keywords were case-normalized to lower-case letters, all stop words were removed, and the Porter stemmer was applied to reduce word variations to their stems or roots [41]. The individually processed bag of title keywords for one user was converted into a title keyword vector that consisted of keywords with term frequency/inverse *user* frequency (TF/IUF) weights. TF/IUF is an adaptation of the original term frequency/inverse *document* frequency (TF/IDF) formula, where each corresponding user is treated as a document [28]. The user frequency indicates how many users have bookmarked references with titles that contain the corresponding keywords.

The second metadata-based similarity is *author-name based similarity*. When a user has bookmarked several papers written by a particular author, we can easily infer that the user's interests reside on the author's research topics. Hence, author names are important anchors to find interesting references and are indicative of interests on related topics. In each user's bookmark repository, all author names of their bookmarked references were aggregated into one bag. The author-name bag for each user was separately converted into an author-name vector with the TF/IUF values. Since author names are proper nouns, text-processing techniques were not applied. The user frequency in the author-name vector counts the number of users who bookmarked papers written by a specific author.

The last metadata-based similarity measure is similarity based on users' social tags. Since social tags provided by a user represent their conceptual understanding or categorization of a reference [33], the tags can serve as rich evidence of user interests. When two users have many common tags, it could be considered as evidence that their interests are similar to each other. In this study, tag-based similarity is calculated based on bi-partite (user-tag) relations (namely, how frequently two users apply the same tags regardless of the annotated items). Before building a social tag vector, we cleaned up pseudo-tags such as "no-tag" and "*file-import-xx-xx-xx." Because the pseudo-tags are automatically assigned by the CiteULike system, they do not carry any information about user interests. Next, we made a tag bag that consisted of a user's whole system of tags. The same text-processing techniques used for paper titles (case normalization, stop-word removal, and stemming) were applied to the individual tag bags. For the tag vector, we computed the TF/IUF values, where the user frequency counts the number of users who annotated their bookmarks with the corresponding tag.

Once all metadata vectors were built (titles, author names, and tags), the similarity between a pair of users was computed using cosine similarity [29].

4.2. *Information Similarity between Co-Members of the Same Community*

In this section, we explore whether the members of the same community share similar interests by testing the H.1 hypothesis. We do so by computing and comparing all seven kinds of similarity for the pairs with co-memberships of the same group, as well as for the pairs without any social connection.

Table 2 shows the comparison of similarity measures. For four bookmark-based measures, co-member pairs have higher similarity than the pairs of unconnected users. All differences are statistically significant (t = 32.22, p <.001 for the number of co-bookmarks; t = 58.62, p < .001 for the Jaccard coefficient; t = 32. 24, p <.001 for the log-likelihood similarity; and t = 54.54, p <.001 for the popularity measure).

Table 2. Similarity of group co-members

	Bookmark-based Similarity				Metadata-based Similarity		
	No. of Co-bookmarks	**Jaccard**	**Log-Likelihood**	**Popularity**	**Title**	**Author Name**	**Social Tag**
Group Co-Members	.26	1.59%	5.0%	8.00	11.2%	2.2%	6.0%
User Pairs Not Socially Connected	.04	0.02%	2.3%	6.92	1.5%	0.1%	0.2%

The results of metadata-based similarity comparison (refer to the last three columns) demonstrate the same patterns. For three metadata-based measures, group co-members had significantly higher similarities than the non-connected pairs (t = 53.69, p <.001 for title-based similarity; t = 322.47, p < .001 for author-name based similarity; t = 165.71, p <.001 for tag-based similarity). As a result, hypothesis 1 can be conclusively accepted for all examined similarity measures; group co-members bookmarked more similar items and shared more comparable topics of interests than users without any group association.

For our next analysis, we examined the usefulness of group libraries. For users engaged in community-based sociality, CiteULike provides access to two kinds of information sources: (1) personal repositories of co-members that we explored above; and (2) group libraries. The question we are addressing in this subsection is whether both kinds of sources are helpful for a user when he is looking for valuable information. To answer this question, we compared which of these two sources holds more items that are similar to a targeted users' information collection.

Before executing this analysis, each member's contributions to his group library need to be excluded. Otherwise, when counting the common items between a member's collection and his group libraries, the items contributed by the member will be mistakenly counted as matches. In addition, for personalized recommendations based on community membership in section 5, when a user's own contributed items are included as a part of candidate items for their recommendations, this is considered a data leakage. It is an unintentional bug to provide ground truth in the training data [18]. As a way to recognize each member's own contributions to their group library without any system log record, we used the timestamp of each bookmark. When one item is posted on both a member's personal repository and their group library with the time lag of not more than 60 seconds, we can assume that the item was posted by that member.[2]

Table 3 shows the results of the comparison. In case of the number of co-bookmarked items, the results of the independent t-test demonstrated that users have significantly more common items with their group libraries (M = 2.63) than with co-members' repositories (M = .26, t = 60.39, p < .001). The popularity-adjusted measure and log-likelihood also shows that users share a much larger fraction of user-specific and rare items with their group libraries than with their co-members' personal repositories (t = 42.44, p < .001 for the popularity weight of co-bookmarked items; t = 103.34, p < .001 for the log-likelihood similarity). However, no significant difference was found for the Jaccard coefficient (t = -1.34, p = 0.18).

---

[2] In general, the presence of the same item in a member's personal repository and their group library can indicate two cases: one is that the member posted the item to the library (refer to Fig. 1) or another is that the member copied it from there. However, it is highly unlikely that a member has copied an item from the group library to the personal repository within 60 seconds of posting this item to the group library by another user; this would require an onerous action of checking the group library's update to reach a decision to copy within 60 seconds.

Table 3. Information similarity: user collections vs. group libraries and user collections vs. co-members' personal repositories

	Group	**Co-member**
No. of Co-bookmarks	2.63	.26
Jaccard	1.5%	1.6%
Log-likelihood	26.9%	5.0%
Popularity	28.10	8.00
Title	19.8%	11.2%
Author Name	5.6%	2.2%
Social Tag	9.8%	6.0%

All metadata-based similarity measures also confirm that users exhibit more common interests with their group libraries than with co-members' personal collections. The titles and author names of users' bookmarked items were significantly more similar to their group libraries (M = .198 for titles; M = .056 for author names) than to their co-members' collections (M = .112, t = 56.78, p < .001 for titles; M = .022, t = 31.13, p < .001 for author names). Users also used more comparable social tags with their group libraries (M = .098) than with their co-members' repositories (M = 0.60, t = 26.02, p < .001).

Summing up the above evaluations, the interests of members participating in the same group are significantly more similar than the interests of users with no social associations. Moreover, group libraries—communal spaces of communities—appear to be more promising information sources than co-members' personal repositories. In the following section, we leverage our findings about the feasibility of group-based sociality as an effective information source in a deeper level by generating personalized recommendations based on users' community memberships, and compare the overall quality to the CF recommendations.

5. Personalized Recommendations Based on Users' Community Membership

The second part of this paper aims to assess the value of social networks established by users' community memberships. This section assesses the hypothesis H.2. We generate personalized recommendations based on users' self-defined community membership and assess the quality of these recommendations by comparing it with CF recommendations, which are based on anonymous peers.

At the beginning of this section, we introduce four types of recommendation algorithms used in our study, including the baseline CF. The next sub-section describes the way to assess the recommendations. It is followed by an analysis to show the comparative advantage of recommendations based on community membership over CF. The last sub-section examines the extent of the value of this community-based sociality for cold-start users.

5.1. Recommendation Algorithms and the Evaluation Approach

In this study, we explored a range of community-based recommendation approaches. They were contrasted against each other and against the CF baseline. Figure 3 shows the overall design space of our community-based recommendation algorithms.

		Computation of Prediction Probability	
		Non-Personalized	Matrix Factorization + Title & Social Tags
Selection of Peers	Anonymous Peers		CF_CW
	Group Members	Community	GMem_CW
	Group Members + Group Libraries		Group_CW

Figure 3. Overall design space of group membership-based recommendation algorithms

5.1.1. Collaborative Filtering Recommendations

Among the various algorithms for CF recommendations [40], we chose the singular vector decomposition (SVD) with alternating least-squares with weighted regularization (ALS-WR) factorizer [5, 8, 11]. The first step of SVD is to compute 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 two f-dimensional lower-rank matrices: $R \approx PQ^T$. One matrix P presents users' latent factor vectors, where $P \in \mathbb{B}^{l \times f}$. The u-th row of P matrix, p_u is a user u's vector, of which elements indicate the degrees of the user's association with f latent factors. Another matrix Q presents items' latent factor vectors, where $Q \in \mathbb{B}^{n \times f}$. The i-th row of Q vector, q_i, is an item i's vector, of which elements represent the degrees of the item's association with f latent factors. The inner product between the two vector elements approximates the probability of the missing bookmarks, $\hat{b}_{ui}$ (i.e. how likely a certain user u is going to bookmark item i) [6] as the following.

$$\hat{b}_{ui} = p_u q_i^T \qquad \text{Eq. (4)}$$

However, the major challenge of SVD is to minimize the error of prediction, $e_{ui} = b_{ui} - \hat{b}_{ui}$. When users bookmarked a tiny percentage of items, the user-item bookmark matrix has many holes (missing values). As a result, the typical SVD algorithm cannot find P and Q. One solution for this problem is to fill out the missing values with some arbitrary numbers, such as zero or one. However, when a system fills out the missing values in the wrong way, even if it is possible to find P and Q, the results are highly prone to overfitting [22]. As the way to solve both data sparsity and overfitting problems in SVD, Zhou et al. [47] proposed the ALS-WR. It aims to learn latent-factor matrices by fitting the existing bookmark records. From known bookmark records, we model P and Q with a weighted-λ-regularization, as in Eq. (5):

$$(P^*, Q^*) = \min_{P^*, Q^*} \sum_{(u,i) \in K} {e_{ui}}^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \qquad \text{Eq. (5)}$$

where K contains all (u, i) pairs for which the b_{ui} is known (the training set). λ is a constant to control the degree of regularization and is usually decided by cross-validation [11]. The next stage is to compute the alternating least squares (ALS). ALS aims to minimize noises and possible errors on the above-regularized model. The basic idea is that when both P and Q are unknown it is possible to fix one of the matrices and to compute another matrix again by solving a least-squares problem. This algorithm alternates the steps by fixing one or both of the matrices. Specifically, 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 input variables and the value of his bookmarks (b_{ui}) as output variables.

$$A_u = Q[u]^T Q[u] = \sum_{i \in B_u} q_i q_i^T \qquad \text{Eq. (6)}$$

$$d_u = Q[u]^T b_u = \sum_{i \in B_u} b_{ui} q_i \quad \text{Eq. (7)}$$

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 value for p_u by the ridge regression, as

$$p_u = (\lambda l_u E + A_u)^{-1} d_u \quad \text{Eq. (8)}$$

$$q_i = (\lambda n_i E + A_i)^{-1} d_i \quad \text{Eq. (9)}$$

E is the f-dimensional identity matrix and $E_f \in \mathbb{B}^{f \times f}$. 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 Eq. (9) (n_i is the number of users who bookmarked item i) [11].

5.1.2. Community Membership-based Recommendations

Our community membership-based recommendation algorithms aim to assess whether community-based sociality possesses a comparable quality to anonymous peers providing CF recommendations. Therefore, we adopted a strategy to replace the anonymous peers with users' community-based sociality. In specific, in selectively choosing community-based social peers, we proposed three strategies: 1) a '*GMem*' approach, to generate recommendations only based on co-members' personal repositories; 2) a '*Group*' approach, to generate recommendations based on both co-members' personal repositories and group library; 3) a '*Community'* approach, to simply pick up the most popular items among co-members without applying any personalization techniques.

Before executing prediction probability computations for every target user, the *GMem* recommendations build a separate submatrix that was only made up of the target user's bookmarks and his group co-members' personal repositories. The second recommendation approach ('Group') uses the submatrix that is composed of a target user's bookmarks, group co-members' personal repositories, and his group libraries (except for his own contributions to group libraries).

When computing prediction probabilities of community-based recommendation approaches, we used the same matrix factorization approach as the CF recommendations. Specifically, the matrix factorization was individually applied to every submatrix. Due to the small number of bookmarks and the denser matrix, the recommendations could be efficiently generated and updated. The standard matrix factorization approach feeds all bookmark data of the whole user population onto the computation at one time, and requires an accordingly significant memory size. Moreover, the CF recommendation algorithm requires a large number of latent factors (for example, over 100 for the Netflix dataset [47]) that are directly correlated to the computational cost. Because the members of a group tend to have a narrow focus on a certain topic, the submatrix-based factorization does not require us to compute a large number of latent factors. In a situation where online users' participation is increasing, the large-scale based CF recommendation approach has a serious scalability problem [35]. In contrast, for our group membership-based recommendations, whenever there are modifications of the bookmarks of a group's social connections, we can simply update the corresponding submatrix and apply the matrix factorization method.

Finally, the community vote approach (marked as *Community*) is the simplest and non-personalized approach among the approaches explored in our study. For each target user, we simply chose the most popular items in his group libraries and co-members' repositories while eliminating the items that the target user had already bookmarked. To determine popularity, we counted how many times each item appeared in group libraries and co-members' collections. Then, the items were ranked in descending order of popularity and were suggested as community vote-based recommendations.

5.1.3. Adding Metadata Information of Items in Recommendations

The results from Section 4 indicate that it is critical to consider not only users' bookmarks, but also the metadata of bookmarked items to represent shared interests. Because the bookmark records do not have rating values, the presence of bookmarks does not carry the reason and the degree of the bookmarkers' interests. Accordingly, the recommendation approaches that are based solely on bookmark records, whether they are based on CF or community-based approaches, merely count the number of co-appearance of items in given users' collections. That means, when some candidate items bookmarked by the same set of peers, the final prediction probability of each candidate item would be exactly the same, even though the items are likely to have different metadata, different content, and probably had different reasons to be bookmarked. Therefore, by using bookmark presence alone, it might be difficult to produce elaborate and accurate suggestions. As a way to enhance the accuracy of recommendations, we propose a strategy to incorporate the metadata information of items in recommendations: content similarity weights (namely, "_CW" postfix).

Specifically, we use the content similarity weights to integrate users' perception of the items (represented by their social tags) and item content (represented by the title keywords) into recommendations. Once a list of candidate items is selected from CF recommendations or community-based recommendations, on the last step to choose the final suggestions shown to the users, the system compares how the metadata of the candidate items—title keywords-based and tag-based similarities—are similar to a target user's keywords and tags. The following equation denotes the way to apply the content similarity weights, for instance, to the CF recommendations.

$$\mathrm{CF\_CW}_{u,i} = \frac{\sum_{v \in p_u} SVD\_Prob_{\ u,v}}{V} \times \left(\left(\frac{k_{ui} - \mu_K}{\sigma_K} \right) + \left(\frac{t_{ui} - \mu_T}{\sigma_T} \right) \right) \qquad \text{Eq. (10)}$$

The above equation shows CF recommendations using the content similarity weight (CF_CW) of a candidate item i for a target user u. For every candidate item, the system calculates the cosine similarity (k_{ui}) between a user u's title keyword vector k_u and the title keyword vector of item i, k_i. It also calculates another cosine similarity (t_{ui}) of a user u's tag-based vector t_u and the tag-based vector of item i, t_i. These two cosine similarities are combined as one content similarity weight. In order to normalize the values, we applied the *standard score* using the mean μ and standard deviation σ of the corresponding similarity measure, respectively [10]. The content similarity value is multiplied by the final prediction probability value of matrix factorization recommendation (namely, the SVD algorithm) for the item i. In this case, v denotes one of the user u's anonymous peers (p_u) who has the candidate item i in his bookmark collection. V is the number of peers who have the candidate item i in their bookmarks. In the same way, the content similarity weights were applied to community-based recommendations, as well.

In our preliminary experiment, we found that all types of recommendations were significantly improved when they were fused with the content similarity weights. Hence, in this study, we do not consider recommendations without content similarity weights.

5.1.4. The Evaluation Approach

To evaluate the proposed recommendation algorithms, we focused on 8,009 target users who have community membership. The target users' bookmark sets were randomly split into 10 equal-sized subsets so as to execute a ten-fold cross-validation. For each iteration, one set of the 10 subsets was used as a test set, and the other nine sets were used as a training set. This process was repeated 10 times with a different test set each time. Recommendations were assessed by the number of *hits*; i.e., the recommended items that are present in the test set. We used traditional evaluation methods for information retrieval, precision, and recall, since the bookmarks in the CiteULike dataset do not have numeric ratings. The measure of precision aims to measure the accuracy of the recommendations, and the measure of recall aims to measure the recommendations' completeness.

Recommendations are usually displayed in a ranked list. Users expected that items in higher ranks would be more important than items in lower ranks. Therefore, the position of correct recommendations is a critical evaluation criterion of recommendations. Specifically, the precision at rank *N* (*precision@N*) is, in the top *N* recommendations, the ratio of correctly predicted items to the number *N* (Eq. 11). The recall at rank *N* (*recall@N*) is, in the top N recommendation, the ratio of correctly predicted items to the total number of relevant items (Eq. 12). We computed the precision and recall, according to three different top *N* ranks (top 10, top 5, and top 2).

$$\text{precision@N} = \frac{\text{No. of correct prediction}}{\text{N of top N set}} = \frac{\text{test} \cap \text{top N}}{\text{N}} \qquad \text{Eq. (11)}$$

$$\text{recall@N} = \frac{\text{No. of correct prediction}}{\text{size of test set}} = \frac{\text{test} \cap \text{top N}}{\text{test}} \qquad \text{Eq. (12)}$$

It is known that the precision and recall move in opposite direction with the growth of N; i.e., the recall is improving while the precision is decreasing. To address this problem, some evaluations prefer to use a single measure, the harmonic mean of precision and recall (F1). However, in this study, in order to examine the different impact of evaluated approaches on both precision and recall, we did not compute the F1 measure.

5.2. The Results of Comparative Evaluation of Community-Based Recommendation Approaches

This subsection attempts to confirm the second hypothesis by providing evidence that community-based sociality is a valuable information source for generating personalized recommendations. To achieve this goal, we compared the quality of three community-based recommendation approaches with the baseline CF recommendations. Figure 4 displays the precision and recall results, respectively. The one-way ANOVA analysis shows that the four recommendation approaches performed differently, with a statistical significance for both evaluation criteria and in all ranks (F = 111.17, p < .001 for top 10 precision; F = 99.34; p < .001 for top 5 precision; F = 71.96, p < .001 for top 2 precision; F = 51.97, p < .001 for top 10 recall; F = 40.20; p < .001 for top 5 recall; F = 24.51, p < .001 for top 2 recall). To analyze the patterns of the differences in details, we executed a post-hoc Schaffé pairwise comparison.

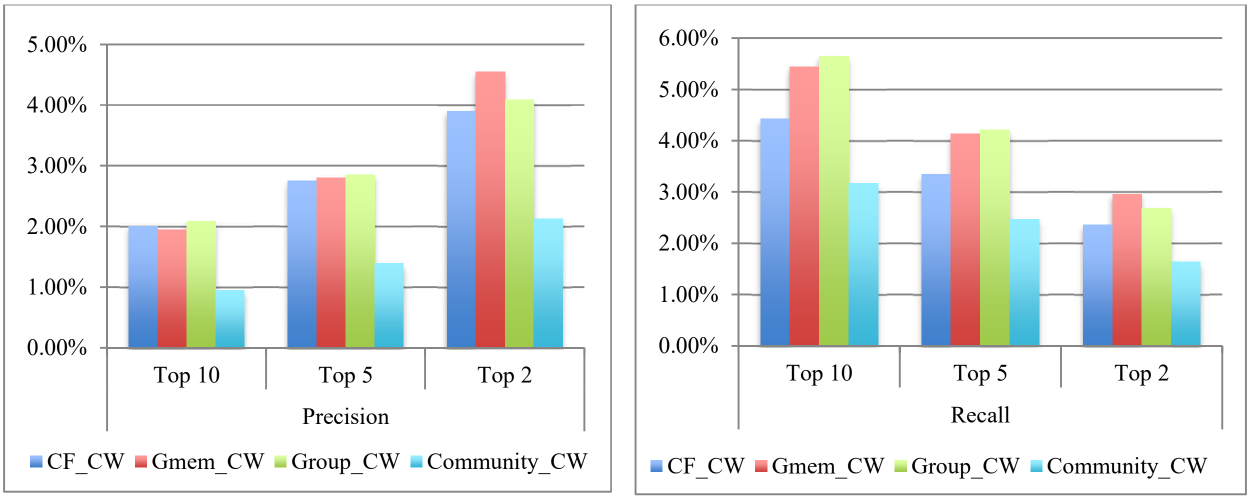

Figure 4. Differences in recommendation quality, depending on the approaches used

First, we tested whether community-based recommendation approaches achieved better quality than baseline CF recommendations. With respect to precision, the recommendations based on users' community membership (either 'GMem_CW' or 'Group_CW') were not significantly more accurate than the CF recommendation approach in all ranks. However, the analysis of the recall demonstrated a different pattern. In all ranks, both GMem_CW and Group_CW significantly outperformed the CF approach. As a result, while hypothesis H.2 cannot be accepted in respect to precision, it can be accepted in respect to recall measures. The community membership-based recommendations can produce suggestions that are as accurate as the baseline CF recommendations, while yielding significantly more complete suggestions than the CF recommendations.

Next, we compared three community-based approaches—GMem_CW, Group_CW, and Community_CW—in order to find the most effective approach using users' community membership. Among them, the non-personalized approach (Community_CW) performed significantly worse than the other two approaches. There was no significant difference between the two remaining approaches, Group_CW and GMem_CW. The quality of recommendations that were produced solely using co-members' repositories was equivalent to the one based on both co-members' repositories and group libraries. The worst performance of the non-personalized community vote shows that although group co-members share similar interests, summing up the shared interests without any personalization is not sufficient. To produce best results and fully leverage the wisdom of the community, more sophisticated community-based recommendations are required, such as GMem_CW and Group_CW.

5.3. Community-Based Recommendations for Cold-Start Users

In our dataset, 2,589 target users (32.33% of our target users) have fewer than five bookmarks in their personal repository. For CF approaches, it is difficult to generate good recommendations for this type of users (known as new users), because their collections are too small to reliably represent their interests. This weakness of the CF recommendations is known as a *cold-start user* problem. Some researchers suggest that social network-based recommendations could be a good solution to solve this cold-start user problem [7, 37]. In line with existing work, this section examines whether a community-based recommendation is also a good solution for the cold-start user problem. We compared the quality of four recommendation algorithms for 2,589 cold-start users.

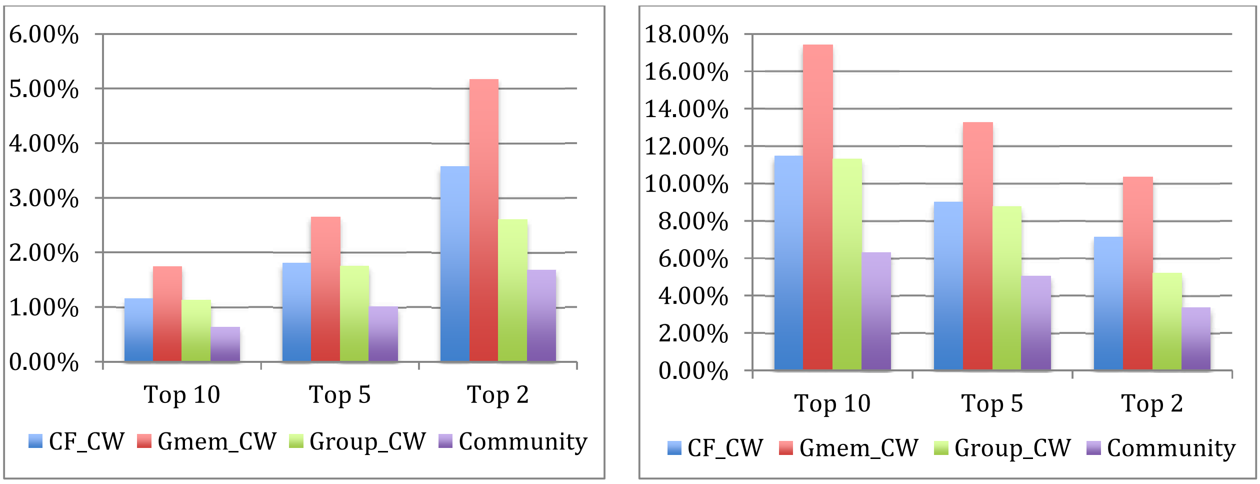

Figure 5. Quality Differences in recommendations for cold-start users

Figures 5 shows the results of precision and recall comparisons, respectively. We investigated the pattern of differences on the various recommendation approaches for new users. We found that both precision and recall tests yielded statistically significant differences ($F = 23.89$, $p < .001$ for top 10 precision; $F = 21.48$, $p < .001$ for top 5 precision; $F = 16.02$, $p < .001$ for top 2 precision; $F = 20.73$, $p < .001$ for top 10 recall; $F = 15.01$, $p < .001$ for top 5 recall; and $F = 10.90$, $p < .001$ for top 2 recall). The recommendations based on group members' personal repositories ('GMem_CW') consistently generated better recommendations and thus outperformed CF recommendations by a large margin in all ranks. In the top 2 results of GMem_CW, there was a 44.81% precision improvement and a 44.67% recall improvement over the CF recommendations. However, unlike similar levels of performance between 'GMem_CW' and 'Group_CW' for overall user population as the Figure 4 depicts, 'Group_CW' did not perform well for cold-start users. The results of the section 4.2 showed that users share user-specific and rate items with their group libraries. On the other hand, co-members' personal repositories may reflect larger variety and multi-topic of interests of individual members. Therefore, for cold-start users who have not expressed their interests sufficiently, co-members' repositories consisting of diverse items are more useful information source than group libraries consisting of items of narrow-topics. The results of this analysis shows that in situations where users' preference data is insufficient, users' self-defined social connections, such as association based on community activities, could be especially valuable as a source of useful information.

6. Conclusion and Discussion

In this paper, we examined the feasibility and the value of user's community membership (in our case, CiteULike groups) for generating personalized recommendations. The first part of this study focused on the feasibility aspect, and demonstrated that the information similarities between users belonging to the same community are significantly higher than the similarities between users in a random (not socially associated) pair. We also discovered that users share more common information with their group libraries than with personal repositories of co-members of the same communities.

The second part of this study focused on the value aspect. By comparing several community-based recommendation approaches, we demonstrated that the recommendations based on community membership produce equivalent or better quality suggestions than CF recommendations. The two advanced community-based recommendations approaches (the one based on community co-members' collections, and another based on both co-members' collections and communities' communal spaces) performed equally well while significantly outperforming a more primitive "community-vote" approach. The critical point here is that community-based recommendations exploited a submatrix that consists of a much smaller number of peers than the CF recommendations. That is, with fewer peers who were explicitly defined by users, our community-based approach produced equivalently accurate and more complete recommendations than the CF recommendation approach.

At the same time, when we separately assessed the performance of recommendation approaches for cold-start users, the advanced community-based approach based on community co-members' collections performed exceptionally well, and outperformed other algorithms by a large margin. This result demonstrates that community wisdom encapsulated in community-based information acquisition is especially valuable for generating recommendations for users who need information aids.

Regarding the merit of communal spaces as a useful information source, the first feasibility part produced supporting evidence that users generally shared more similar information with their group libraries than with co-members' personal repositories. However, the second part failed to show any additional value of group libraries for generating recommendations. The recommendations based only on co-members' repositories were either equally good or better than the ones based on both co-members' repositories and group libraries. We speculate that the nature of recommendation algorithms may account for the result. In the first part, we made a comparison between all group libraries and all of each user's co-members. However, personalized recommendations must pin down a few of the most like-minded social peers and choose favorable items in the peers' collections.

Altogether, this study presents three important specific contributions. First, this study empirically demonstrates the shared interests among members of the same community. Second, we suggest several recommendation approaches that are based on community membership, and demonstrate their benefits over traditional collaborative filtering. Third, we demonstrate that community-based sociality is an especially valuable source of recommendations for cold-start users. This study could be considered as one of the few attempts to examine the feasibility and substantiate the practical value of community-based social networks as a source of useful information for recommendations. The results of this study cast light on how to adequately provide personalized recommendations for individual users by using their online community membership information.

It is important to acknowledge that our study has several limitations, which we will try to address in future work. While our study did help us to determine the best approaches among those we explored (in particular, outperforming the state-of-the-art CF approach), the precision and recall values of all approaches examined in this study are not high enough. We hope to find a better way to improve the overall quality of recommendations. We believe that the main problem with the quality of recommendations is the weakness of bookmarks as evidence of user interests. Since the bookmarking process is somewhat opportunistic, the *presence* of a bookmark certainly indicates that users have interests in certain items, but an *absence* of a bookmark does not necessarily indicate a total lack of interest. We plan to explore more sophisticated approaches to combine bookmarking and metadata sources to increase future performance. We also intend to explore a broader set of metadata; as previously mentioned, some metadata available in CiteULike were left out because they were available only for a subset of the papers. Finally, this study focuses on one system: CiteULike. To examine the generality of our results, we will explore other systems with community-based social connections that focus on different domains with different types of target items and different types of metadata. This research will show whether our feasibility hypothesis will hold for other domains, as well as to what extent the proposed recommendation approaches are applicable in other social systems and domains.

Acknowledgment

We are grateful to Professor Ian Ruthven and anonymous reviewers for their constructive comments, which helped to improve the manuscript.

Reference

[1] Adamopoulos, J. (2012). The emergence of social meaning: A theory of action construal *Handbook of Social Resource Theory* (pp. 255-272): Springer.

[2] Bakhshi, S., Kanuparthy, P., & Gilbert, E. (2014). *Demographics, weather and online reviews: a study of restaurant recommendations*. Proceedings of the 23rd international conference on World wide web, Seoul, Korea.

[3] Bello, G. A., Harenberg, S., Agrawal, A., & Samatova, N. F. (2016). Community Detection in Dynamic Attributed Graphs. In J. Li, X. Li, S. Wang, J. Li, & Q. Z. Sheng (Eds.), *Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings* (pp. 329-344). Cham: Springer International Publishing.

[4] Cantador, I., & Castells, P. (2012). Group Recommender Systems: New Perspectives in the Social Web *Recommender Systems for the Social Web* (Vol. 32, pp. 139-157): Springer Berlin Heidelberg.

[5] Chan, S., & Treleaven, P. (2015). Continuous Model Selection for Large-Scale Recommender Systems. *Big Data Analytics, 33*, 107.

[6] Chui, C. K., & Chen, G. (2012). *Signal processing and systems theory: selected topics* (Vol. 26): Springer Science & Business Media.

[7] Deng, S., Huang, L., Xu, G., Wu, X., & Wu, Z. (2016). On Deep Learning for Trust-Aware Recommendations in Social Networks. *IEEE Transactions on Neural Networks and Learning Systems, PP*(99), 1-14.

[8] Dhumal, P., & Deshmukh, S. (2016). Retrieval and extraction of unique patterns from compressed text data using the SVD technique on Hadoop Apache MAHOUT framework. Proceedings of 2016 International Conference on Computing Communication Control and automation (ICCUBEA),.

[9] Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading collaboration in online communities. *Mis Quarterly, 39*(2), 393-412.

[10] Fiasconaro, A., Tumminello, M., Nicosia, V., Latora, V., & Mantegna, R. N. (2015). Hybrid recommendation methods in complex networks. *Physical Review E, 92*(1), 012811.

[11] Friedman, A., Berkovsky, S., & Kaafar, M. A. (2016). A differential privacy framework for matrix factorization recommender systems. *User Modeling and User-Adapted Interaction, 26*(5), 425-458.

[12] Garcia, I., Sebastia, L., & Onaindia, E. (2011). On the design of individual and group recommender systems for tourism. *Expert Systems with Applications, 38*(6), 7683-7692.

[13] Gras, B., Brun, A., & Boyer, A. (2016). *Identifying Grey Sheep Users in Collaborative Filtering: A Distribution-Based Technique*. Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, Halifax, Nova Scotia, Canada.

[14] Gunes, I., Kaleli, C., Bilge, A., & Polat, H. (2014). Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review, 42*(4), 767-799. doi:10.1007/s10462-012-9364-9

[15] Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge Sharing in Online Communities: Learning to Cross Geographic and Hierarchical Boundaries. *Organization Science, 26*(6), 1593-1611.

[16] Jonnalagedda, N., Gauch, S., Labille, K., & Alfarhood, S. (2016). Incorporating popularity in a personalized news recommender system. *PeerJ Computer Science, 2*, e63.

[17] Kaššák, O., Kompan, M., & Bieliková, M. (2016). Personalized hybrid recommendation for group of users: Top-N multimedia recommender. *Information Processing & Management, 52*(3), 459-477.

[18] Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data, 6*(4), 1-21. doi:10.1145/2382577.2382579

[19] Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education, 56*(3), 885-899.

[20] Kompan, M., & Bielikova, M. (2013). Personalized recommendation for individual users based on the group recommendation principles. *Studies in Informatics and Control, 22*(3), 331-342.

[21] Kompan, M., & Bielikova, M. (2014). Group recommendations: Survey and perspectives. *Computing and Informatics, 33*(2), 446-476.

[22] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer, 42*(8), 30-37. doi:10.1109/mc.2009.263

[23] Lee, D., & Brusilovsky, P. (Forthcoming). Social Link-based Recommendations: A Review. In P. Brusilovsky & D. He (Eds.), *Social Information Access*. Heidelberg: Springer.

[24] Lee, D. H. (2013). *Personalized Recommendations Based on Users' Information Centered Social Networks.* (Doctoral thesis Doctoral Dissertation), University of Pittsburgh, Pittsburgh, PA, USA. Retrieved from http://d-scholarship.pitt.edu/18546/

[25] Lou, J., Fang, Y., Lim, K. H., & Peng, J. Z. (2013). Contributing high quantity and quality knowledge to online Q&A communities. *Journal of the American Society for Information Science and Technology, 64*(2), 356-371.

[26] Manca, M., Boratto, L., & Carta, S. (2014). *Mining User Behavior in a Social Bookmarking System-A Delicious Friend Recommender System.* Proceedings of the 3rd International Conference on Data Management Technologies and Applications.

[27] Manchanda, P., Packard, G., & Pattabhiramaiah, A. (2015). Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community. *Marketing Science, 34*(3), 367-387.

[28] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*: Cambridge University Press.

[29] Meng, L., Huang, R., & Gu, J. (2013). A review of semantic similarity measures in wordnet. *International Journal of Hybrid Information Technology, 6*(1), 1-12.

[30] Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research, 15*(4), e85.

[31] Mori, J., Tsujishita, T., Matsuo, Y., & Ishizuka, M. (2006). Extracting Relations in Social Networks from the Web Using Similarity Between Collective Contexts *Proceedings of 5th International Semantic Web Conference on the Semantic Web - ISWC 2006* (Vol. 4273, pp. 487-500): Springer Berlin Heidelberg.

[32] Nicolini, D., Mengis, J., & Swan, J. (2012). Understanding the role of objects in cross-disciplinary collaboration. *Organization Science, 23*(3), 612-629.

[33] Raban, D. R., Danan, A., Ronen, I., & Guy, I. (2016). Impression management through people tagging in the enterprise: Implications for social media sampling and design. *Journal of Information Science*.

[34] Ram Naresh Yadav, B. V., Satyanarayana, B., & Vasumathi, D. (2016). A Vector Space Model Approach for Web Attack Classification Using Machine Learning Technique. In C. S. Satapathy, S. K. Raju, K. J. Mandal, & V. Bhateja (Eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies: IC3T 2015, Volume 3* (pp. 363-373). New Delhi: Springer India.

[35] Rao, N., Yu, H.-F., Ravikumar, P. K., & Dhillon, I. S. (2015). *Collaborative filtering with graph information: Consistency and scalable methods.* Advances in Neural Information Processing Systems.

[36] Razmerita, L., Kirchner, K., & Nabeth, T. (2014). Social Media in Organizations: Leveraging Personal and Collective Knowledge Processes. *Journal of Organizational Computing and Electronic Commerce, 24*(1), 74-93.

[37] Sedhain, S., Sanner, S., Braziunas, D., Xie, L., & Christensen, J. (2014). *Social collaborative filtering for cold-start recommendations*. Proceedings of the 8th ACM Conference on Recommender systems, Foster City, Silicon Valley, California, USA.

[38] Sheng, M., & Hartono, R. (2015). An exploratory study of knowledge creation and sharing in online community: a social capital perspective. *Total Quality Management & Business Excellence, 26*(1-2), 93-107.

[39] Steck, H. (2011). *Item popularity and recommendation accuracy*. Proceedings of the fifth ACM conference on Recommender systems, Chicago, Illinois, USA.

[40] Su, X. (2015). *Collaborative filtering: A survey.* Proceedings of 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions).

[41] Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management, 50*(1), 104-112.

[42] Wang, H., Li, G., & Feng, J. (2014). Group-Based Personalized Location Recommendation on Social Networks. In L. Chen, Y. Jia, T. Sellis, & G. Liu (Eds.), *Web Technologies and Applications* (Vol. 8709, pp. 68-80): Springer International Publishing.

[43] Yuan, Q., Chen, L., & Zhao, S. (2012). Augmenting Collaborative Recommenders by Fusing Social Relationships: Membership and Friendship *Recommender Systems for the Social Web* (Vol. 32, pp. 159-175): Springer Berlin Heidelberg.

[44] Zhang, F., & Zhou, Q. (2014). HHT–SVM: An online method for detecting profile injection attacks in collaborative recommender systems. *Knowledge-Based Systems, 65*, 96-105.

[45] Zhang, H., Shen, F., Liu, W., He, X., Luan, H., & Chua, T.-S. (2016). *Discrete Collaborative Filtering*. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, Pisa, Italy.

[46] Zhao, Y.-L., Nie, L., Wang, X., & Chua, T.-S. (2014). Personalized Recommendations of Locally Interesting Venues to Tourists via Cross-Region Community Matching. *ACM Trans. Intell. Syst. Technol., 5*(3), 1-26. doi:10.1145/2532439

[47] Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). *Large-Scale Parallel Collaborative Filtering for the Netflix Prize*. Proceedings of the 4th international conference on Algorithmic Aspects in Information and Management, Shanghai, China.

[48] Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). Large-Scale Parallel Collaborative Filtering for the Netflix Prize. In R. Fleischer & J. Xu (Eds.), *Proceedings of the Algorithmic Aspects in Information and Management: 4th International Conference, AAIM 2008, Shanghai, China, June 23-25, 2008.* (pp. 337-348). Berlin, Heidelberg: Springer Berlin Heidelberg.