IMPROVING COLLABORATIVE FILTERING RECOMMENDER BY USING MULTI-CRITERIA RATING AND IMPLICIT SOCIAL NETWORKS TO RECOMMEND RESEARCH PAPERS

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Shaikhah Alotaibi

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science

176 Thorvaldson Building

110 Science Place

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5C9

Abstract

Research paper recommender systems (RSs) aim to alleviate the information overload of researchers by suggesting relevant and useful papers. The collaborative filtering in the area of recommending research papers can benefit by using richer user feedback data through multi-criteria rating, and by integrating richer social network data into the recommender algorithm.

Existing approaches using collaborative filtering or hybrid approaches typically allow only one rating criterion (overall liking) for users to evaluate papers. We conducted a qualitative study using focus group to explore the most important criteria for rating research papers that can be used to control the paper recommendation by enabling users to set the weight for each criterion. We investigated also the effect of using different rating criteria on the user interface design and how the user can control the weight of the criteria. We followed that by a quantitative study using a questionnaire to validate our findings from the focus group and to find if the chosen criteria are domain independent.

Combining social network information with collaborative filtering recommendation algorithms has successfully reduced some of the drawbacks of collaborative filtering and increased the accuracy of recommendations. All existing recommendation approaches that combine social network information with collaborative filtering in this domain have used explicit social relations that are initiated by users (e.g. “friendship”, “following”). The results have shown that the recommendations produced using explicit social relations cannot compete with traditional collaborative filtering and suffer from the low user coverage. We argue that the available data in social bookmarking Web sites can be exploited to connect similar users using implicit social connections based on their bookmarking behavior. We explore the implicit social relations between users in social bookmarking Web sites (such as CiteULike and Mendeley), and propose three different implicit social networks to recommend relevant papers to users: readership, co-readership and tag-based implicit social networks. First, for each network, we tested the interest similarities of users who are connected using the proposed implicit social networks and compare them with the interest similarities using two explicit social networks: co-authorship and friendship. We found that the readership implicit social network connects users with more

similarities than users who are connected using co-authorship and friendship explicit social networks. Then, we compare the recommendation using three different recommendation approaches and implicit social network alone with the recommendation using implicit and explicit social network. We found that fusing recommendation from implicit and explicit social networks can increase the prediction accuracy, and user coverage. The trade-off between the prediction accuracy and diversity was also studied with different social distances between users. The results showed that the diversity of the recommended list increases with the increase of social distance.

To summarize, the main contributions of this dissertation to the area of research paper recommendation are two-fold. It is the first to explore the use of multi-criteria rating for research papers. Secondly, it proposes and evaluates a novel approach to improve collaborative filtering in both *prediction accuracy* (performance) and *user coverage and diversity* (nonperformance measures) in social bookmarking systems for sharing research papers, by defining and exploiting several implicit social networks from usage data that is widely available.

Acknowledgements

This work could not be done without the help, support and encouragement of some people. Firstly, I would like to express my sincere appreciation and gratitude to my supervisor Prof. Julita Vassileva for the continuous support of my Ph.D study and related research, for her patience, motivation, and knowledge. Her guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Gordon McCalla, Prof. Jim Greer and Prof. Chris Zhang, for their insightful and constructive criticism and comments, but also for their questions which convinced me to widen my research from various perspectives. I also would like to thank Dr. Nava Tintarev for agreeing to be my external examiner and volunteering her time to read my thesis in very short time, and providing me with very detailed annotated copy of my thesis which enable me to polish the final version. I also would like to thank Dr. Mark Eramian for chairing the thesis defence.

One of the people who helped me along the way is Gwen Lancaster, thanks Gwen for understanding my needs and fulfill them very quickly. I am also thankful to MADMUC lab members for their friendship and their valuable discussions and opinions.

Thanks from the bottom of my heart to my father, Nayef, and my mother, Norah, for their love and support and endless prayers that I felt every day during my PhD journey. I am also grateful to the most important person in my life, my wonderful husband, Fahad, who sacrificed the most and supported me a lot. He did an amazing job during my absence from my house due to my working on my thesis. Thanks also for my amazing son, Fayez, and my sweet daughters, Norah, Danah and Savannah for being such marvelous children who gave me the happiness and helped me to forget the tiredness. I am also thankful to my brothers and sisters for their continuous love, care and support.

Finally, I would like to thank the Institute of Public Administration, Canadian Government NSERC, and University of Saskatchewan, and for their financial support.

Dedicated:

To Fahad, my love, husband and best friend, I couldn’t done this without you

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

- CF: **C**ollaborative **F**iltering
- CBF: **C**ontent-**B**ased **F**iltering
- RS: **R**ecommender **S**ystem
- SVD: **S**ingular **V**alue **D**ecomposition
- MCRS: **M**ulti-**C**riteria **R**ecommender **S**ystem
- MSVD: **M**ulti-linear **S**ingular **V**alue **D**ecomposition
- SNRS: **S**ocial **N**etwork-based **R**ecommender **S**ystem
- ISN: **I**mplicit **S**ocial **N**etwork
- P@N: **P**recision measure at rank N
- R@: **R**ecall measure at rank N
- F1@N: **F1** measure at rank N
- SR: **S**ocial **R**ecommender
- CR: **C**ombined **R**ecommender
- AR: **A**mplified **R**ecommender
- ANOVA: **AN**alysis **O**f **VA**riance

CHAPTER 1
INTRODUCTION

Scholarly papers fulfil a number of roles: (1) they provide a communications channel for researchers to report their research results; (2) they provide knowledge resources to keep researchers current with new results in their areas of interest; and (3) they serve as directories of other researchers with similar interests with whom researchers could collaborate. However, with the proliferation of publishers, online journals, and conferences, the number of new published papers has become overwhelming. For this reason, many recommender systems (RSs) have been proposed to help readers in these tasks. RSs employ ranking criteria that suggest lists of potential papers to users. The two main algorithms that RSs use are content-based filtering (CBF) and collaborative filtering (CF).

CBF is based on information retrieval techniques that compare a paper's features (e.g., title, abstract, keywords, publication year) with the researchers' features (e.g., interests or previous search queries) to find matches [1]. In contrast, CF (e.g.,[2][3]) does not need domain knowledge to recommend papers, and thus it is the most widely used; CF uses the similarities between previous item ratings to find users who are similar to the target user and then to recommend items that these same users have liked. Hybrid recommendations (e.g., [4]) combine CBF and CF to alleviate the drawbacks of both. The two main drawbacks of CBF are overspecialization (filter bubble) and domain dependency, and one main challenge of CF is data sparsity, which occurs when the rating data are insufficient for identifying similar users. The other important problem is so called "cold start", which occur when the system has just started and there are insufficient ratings for items or users; ratings information is important for identifying the similarities between users and/or items. In addition, most CF algorithms require users to give just one overall (global) rating and then use the averages of all users' ratings to correlate the items (or users) and compute "neighbourhoods." This approach is straightforward but not sufficiently flexible to provide adequate details about the quality of the rated item or service. The inflexibility of global ratings produces biased recommendations because two users may give the same global rating from two different perspectives [5]. For example, two researchers may rate a paper the same, but the first researcher's evaluation may be based on the paper's readability, whereas the other's is based on the paper's novelty. For this reason, some RSs are based on CF algorithms or hybrid approaches that use multi-criteria ratings based on two or more perspectives (dimensions). However, multi-criteria RSs that consider users' subjective opinions are rarely found, and this is a research area that has remained relatively unexplored [6]. In addition, in the existing multi-criteria RSs, researchers choose the rating criteria. To the best of our knowledge, there is only one study that collected users' opinions about the criteria that should be included in rating items or services. However, this study investigated the multi-criteria ratings in the movies domain [7]. Thus, the results cannot be reused in recommending research papers. Unlike in taste domains, such as recommending music, movies, or restaurants, in research paper recommendations the important rating criteria are not clear. Thus, this dissertation investigates users' opinions about multi-criteria rating systems specifically for recommending research papers, and it gathers user requirements for building such a system.

Another way to overcome one or more of the above CF drawbacks is to exploit users' social ties in what is called social network-based recommender system. With the increasing number of social networks in applications such as social bookmarking systems (e.g., CiteULike, Mendeley, ResearchGate, Academia.edu, Delicious), which researchers often use to manage their digital paper and bookmark libraries, users can be connected through different social relations. The idea is to exploit users' social ties in the recommendation process because people generally trust recommendations from friends more than those from unknown others. This phenomenon is explained by homophily theory [8] and social influence theory [9]; homophily theory suggests that users like to be connected to users with similar preferences, whereas social influence theory suggests that users' behaviours are influenced by the behaviours of their connected users. Therefore, by knowing that two users are connected, one can infer that they possibly share interests (which helps in addressing data sparsity and cold starts) and then recommend items from connected users (which helps solve the trust problem).

Following homophily theory, we attempt to find algorithms to construct social networks that connect each user to the most similar users; then we use the social relationships to recommend research papers and people. Because the objective is to find relevant information rather than friends, users may not necessarily be aware that they are connected, just as is the case with similar CF users who are not aware of one another's existence. Rather, the CF recommender system uses similar ratings histories to correlate these users, thus creating virtual, implicit relationships between them. Social bookmarking services rarely have sufficient rating data but instead will provide many other clues to similarities in user interests based on users' behaviours in the system and their paper authorship. However, surprisingly, none of the popular social bookmarking tools have utilized the wealth of social data they store to build a hybrid (i.e., social combined with CF) recommender system. The CiteULike and Mendeley social bookmarking systems provide RSs for their users, but they are not social RSs.

Although the approaches proposed in [10][11]), and [12] provide a good beginning, they have limitations. For example, in the watching network [12], in which users initiate the connections, users who are not watching any other users are not part of any network, and the approach cannot improve their recommendations or address the cold start problem. In a collaboration network [10], users must have publications to have social connections, and new researchers without publications (e.g., students) are excluded. Similarly, new users who have not joined groups are not part of group networks [11].

Thus, recommendations that are based on the relationships initiated by users (i.e., explicit) suffer from low user coverage [10]. The main focus in most studies about recommending research papers specifically and about recommendations in general is on developing algorithms that increase recommendation performance without paying attention to nonperformance measures such as user coverage or diversity of recommendation lists, but these two measures specifically are important to consider. Some recommenders cannot provide recommendations for some users due to the low confidence in prediction accuracy, but it is preferred to provide recommendations to wider ranges of users. Among these recommenders, evaluations should consider the trade-off between prediction accuracy and coverage [13]. However, thinking only about increasing prediction accuracy might result in recommendation lists that have items that are very similar to each other because the focus is to find the items that are most similar to the user's profile; users in some cases want to be exposed to items that are relevant to their research interests but that discuss the similar topics from different perspectives.

In this work, different social networks are built implicitly based on users' bookmarking behaviours with the aim of giving users equal chances to connect to one or more social networks, which would then increase their chances of gathering recommendations (i.e., increasing user coverage). Users with or without publications can be connected using our proposed implicit social networks, which means that users with different levels of expertise are considered when the implicit social networks are proposed.

1.1 Objectives and Research Questions

The research objectives of this dissertation are:

- Understanding user behaviour in evaluating the quality of research papers
- Understanding user perceptions about the importance of multi-criteria rating systems
- Collecting user opinions about the most important rating criteria for research papers and exploring whether these criteria are domain-dependent
- Finding alternative social connections between users other than explicit social connections, which are usually few per user on social bookmarking websites
- Finding recommendation approaches that strike a balance between the *prediction accuracy* and the *user coverage* of recommendations
- Finding recommendation approaches that strike a balance between the *prediction accuracy* and the *diversity* of recommendations

1.1.1 Research Questions Related to Multi-Criteria Rating RSs

Regarding multi-criteria rating RSs, we aim to answer the following questions:

RQ1: How do users perceive multi-criteria rating recommendations?

This broad question is divided into the following specific research questions:

RQ1.1: What are the most important rating criteria in evaluating a research paper?

RQ1.2: What are users' preferences about using overall ratings versus multi-criteria ratings?

RQ1.3: Do users prefer to have control over the importance weights of multi-criteria ratings during the recommendation process?

RQ1.4: Are the criteria domain-dependent?

1.1.2 Research Questions Related to Implicit Social Networks (ISNs)

The part of this dissertation that is based on social recommendations consists of two parts: testing the interest similarity between users in each implicit social network and testing the performance of recommendations that use different sources of information (implicit, explicit, or both). The first part aims to test the interest similarities among users who are socially connected by one of the proposed implicit social networks; we test for these similarities using different similarity measures to investigate whether there are any differences in the results. Because we used bookmarking data when there were no numeric ratings for research papers, there are specific measures that we could use to test the interest similarities for unary data, although there was no evidence of whether one of the measures outperformed the others.

For the first part of the social recommendations aspect of this dissertation, the following questions need to be answered:

RQ2: Comparing three implicit social networks, readership (which consists of one of two types of relationships: reciprocal or unidirectional), co-readership, and tag-based, which one connects the most similar users?

RQ3: In each of the three proposed ISNs, how does the relationship distance between two connected users affect the users' interest similarities?

RQ4: Is the interest similarity between users who are implicitly socially connected comparable to the one between users who are explicitly connected?

For the second part of this aspect of the dissertation, which tests the results of using ISNs as recommendation resources, we aim to answer the following questions:

RQ5*:* What is the effect of using implicit social networks in improving recommendations?

RQ5.1: Using different social recommendation approaches, which approach works the best for each ISN?

RQ5.2: Comparing the recommendations using different ISNs, which one produces the highest prediction accuracy?

RQ5.3: Does fusing recommendations from ISNs and explicit SNs improve the performance of the recommendation?

RQ5.4: What is the effect of social relationship distance on prediction accuracy and the diversity of recommendation lists?

1.2 Research Contributions

The main goal of this dissertation is to alleviate some of the drawbacks of the conventional collaborative filtering approach, and it contributes in different research areas. The main contribution is in the area of recommender systems in general but, more specifically, multi-criteria rating recommender systems and social recommendations with a focus on recommending research papers. This dissertation also contributes to the fields of user modelling, personalization, and adaptation, and it has an impact on lifelong learning given that the items under consideration are research papers, which are mainly the learning objects for researchers. Some parts of this dissertation suggest ideas that enhance recommender system user interfaces, so that it also contributes in the human-computer interaction area.

This dissertation addresses three main points. First, it addresses the lack of research in the field of multi-criteria rating RSs as a whole, and it contributes specifically in multi-criteria rating RSs for recommending research papers. This is the first study to consider user opinions about the multi-criteria ratings of research papers. Unlike systems in the taste domain, such as music or movies, there is no available system with numeric ratings for evaluating the quality of research papers.

Second, three implicit social networks are built based on the publicly available data in social bookmarking websites (e.g., CiteULike): readership, co-readership, and tag-based networks. The objective of building nonexistent (virtual) ISNs is to explore them in order to find better information resources for personalized recommendations rather than explicit social networks that suffer from low coverage. The interest similarities of the three networks are compared with each other, and each network has three variations based on the distance between each two connected users (direct, one-hop, and two-hop relationships). A comprehensive and detailed analysis of the three proposed networks is one of the main contributions of this research, as is understanding the relationships between users and their similarities in terms of helping to build better RSs. One of the main contributions of the dissertation is the proposed of fusing recommendations based on explicit social networks with recommendation based on implicit social networks. To make the recommendation more personalized, the weights that are used for recommendation sources are set based on the contribution of the user in both sources. For example, if the user has relations in only explicit social network, the total weight is set to recommendation based on explicit social network and the implicit social network is ignored.

Third, the dissertation explores the trade-offs between performance and nonperformance measures. The majority of the work on recommender systems focuses on performance measures. However, there are trade-offs between different measures [14]; [13] that make it necessary to study the effect of both measures rather than study the effect of each measure in the isolation of the other. The dissertation first explores the trade-off between prediction accuracy and user coverage; the performance of the recommendations that use social relations that are defined implicitly between users is compared with the performance and nonperformance measures when the recommendations from explicit and implicit social relations are fused. Then, the dissertation explores the trade-offs between prediction accuracy and the diversity of recommendation lists. Because social networks allow for propagating the relationships between users, the roles of distant users are studied in terms of the effects of prediction accuracy and recommendation list diversity.

1.3 Definitions of Terms

- **Explicit social networks:** The social networks in which the user relationships represent undirected real-world relationships that are initiated by users, and users are generally aware of them. These relationships could be based on agreements between any two users, such as being friends on Facebook or connections on CiteULike or LinkedIn; both users know that they are connected and the relationship is by invitation and acceptance. The other kind of explicit relationships is unidirectional, in which one user initiates the connection to the other, and the other user might or might not know or care about identifying the social connection. Examples of these relationships include following people on Twitter or Instagram, watching relationships on CiteULike, or “Liking” things on Facebook.
- **Implicit social network:** Any network in which the relationships between users are inferred from the users’ behaviour patterns, such as co-tagging, co-purchasing, or co-bookmarking an item or co-commenting on the same entry (e.g., blog entry, picture, or

video). In our work, we extend the set of patterns to reflect specific behaviours in the domain of research paper recommendations (co-reading, co-authorship, readership).

- **Social recommendation:** As proposed here using the “narrow” definition [15], any recommendation that includes social relations as extra input to improve it.
- **Social peer (social friend):** Any user with whom the target user has a connection in the social network.
- **Multi-criteria ratings:** When an item (in our case, research papers) has more than one overall rating and each rating represents the user’s opinion about a specific feature of the item.
- **User coverage:** in this dissertation, the user coverage means how many users gets nonempty recommended list from the recommender.
- **Diversity of the recommended list:** The list of diverse items is equal to the list of less similar items.

1.4 Dissertation Outline

The rest of this dissertation is organized as follows:

- **Chapter 2** is dedicated to a literature review on recommender systems, including the main approaches, multi-criteria RSs, social network-based RSs, and RSs for researchers’ use.
- **Chapter 3** contains the details of the qualitative experiment that was conducted, using a focus group to gather information about the most important criteria that users consider in rating a paper followed by a quantitative study to validate our findings from the qualitative study.
- **Chapter 4** explains how the three proposed implicit social networks are constructed, the dataset that is used, and the experiments that are conducted to compare the users’ interest similarities in the three ISNs. The interest similarities in these networks are also compared with two explicit social networks.
- **Chapter 5** discusses the performance of different recommendation approaches considering different ISNs as information sources. In addition, the effect of fusing different recommendations using implicit and explicit social networks is also evaluated in terms of prediction accuracy and user coverage.

- **Chapter 6** examines the effect of social relationship distance on prediction accuracy and diversity of recommendation lists.
- **Chapter 7** is dedicated to the summary of findings and their implications, as well as to conclusions, limitations, and potential future work.

CHAPTER 2:
RELATED WORK

The work in this research tackled different topics. In this section, the related previous work on these topics is discussed. The topics are recommender systems, social recommendations and multi-criteria rating recommendation and scholarly paper recommendations.

2.1 Recommender Systems

Recommender systems (RSs) can be defined as “any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options” (Burke, 2002, p. 331). In this definition, Burke emphasizes two RS criteria that differ from those of information retrieval systems or search engines: “individualized” and “interesting or useful.”

Recommender systems (RSs) are supporting systems that find information, products, services, or people by analyzing their attributes and the reviews given to them by other users to customize the recommendations for the active user who has special tastes, preferences, and needs. Personalization and the adaptation of the recommendation are the most important features of RSs. Today, RSs have become an important part of most of websites. Many are applied in e-commerce applications (i.e., Amazon.com, eBay.com). RSs help people to make decisions in their daily lives, such as in buying items, reading news, or watching movies. RSs are particularly useful in environments where the amount of information is huge and it is difficult for users to make the right decisions. In the case of getting good recommendations, users can save their money, effort, and time.

RSs have gained the attention of researchers in the last two decades. Many RS approaches have been deployed in different application areas including e-commerce, e-health, and e-learning, either as desktop applications or mobile applications. Recommendations could be produced using many approaches, the traditional approaches are: content-based filtering, collaborative filtering, or hybrid approaches, which are combinations different filtering approaches.

2.1.1 Content-based Filtering

Content-based filtering (CBF)—also called "item-to-item correlation" by Schafer, Konstan, & Riedi (1999)—is based on information retrieval techniques that compare and calculate the similarity between an item's features (item profile) with the user's features (user profile), and that show his/her interests or previous search queries when these are available to find matches; therefore, enough information about the user should be available beforehand. The user can feed the information into the system *explicitly* during the registration phase, or it can be deduced by the system *implicitly* from analyzing user behavior such as purchasing or browsing history, downloading, or document printing. The best matching items that are unknown to the user are then recommended. Usually weights are assigned to the attributes to represent their importance. NewsWeeder [17], a filtering system to filter newsgroups, uses the words of the text as features of the news article. Most of the CBF techniques are based on the finding and analyzing similarities between texts. However, purely content (text)-related features may not be sufficient to generate appropriate recommendations. For example, the recommended item could be relevant to the interests of the user based on the attributes used, but the system cannot discover quality issues such as if the item is really worth buying or if the music file quality is high. In addition, content-based approaches can overspecialize the recommendations, causing a "filter bubble," which means users would not have the opportunity to see other items that s(he) may like if these items are not among user current interests (i.e., similar to the ones that the user has rated before).

2.1.2 Collaborative Filtering

Collaborative filtering (CF) is a term first used when Tapestry was developed by Goldberg, Nichols, Oki, & Terry (1992), an experimental e-mail system that filtered documents received by e-mail. CF is the most familiar and the most used filtering technique. In contrast to content-based filtering, which relates item to item, CF relates people to people [18]. CF is helpful if there are few or no features available about the items or the user but user ratings are available. It allows for predicting a user's rating of a new item by correlating his or her previous ratings with the ratings of the same items given by other users and finding a "neighborhood" of similar-minded users. Some advantages of the CF approach are that there is no need for content analysis (which is needed in the content-based approach). It also provides users with serendipitous items that surprise the user and are less similar to the items that the user likes. Serendipity is used to alleviate the effect of overspecialization, which appears in the content-based approach. Another advantage of CF is that its approaches are domain-independent because they are not looking at the item's attributes; CF instead looks to the item ratings, which are not related to a specific domain. Examples of CF systems are Tapestry [19], filters incoming e-mail streams, and GroupLens [20], which filters netNews.

CF approaches often depend user rating can be classified along two dimensions: explicit vs. implicit and overall versus multi-criteria.

- **Explicit ratings versus implicit ratings**

User ratings can be collected explicitly or implicitly. With explicit ratings, the user rates items or services by, for example, indicating like/unlike (i.e., thumb up/thumb down) or by giving a number of stars or a numeric value such as one a Likert scale by which the user can use values from 1 to 5, 1 meaning the worst and 5 meaning the best. Examples of explicit ratings are the ratings used in CiteUlike.com and Amazon.com. Explicit ratings have many advantages such as when the user rating can lead to increased recommendation accuracy. In addition, explicit ratings can be considered as a good resource for historical user preferences. Furthermore, explicit ratings can increase satisfaction because the user feels that his/her opinion has value in the community [21]. However, online communities suffer from lurkers (i.e., free riders) who consume other's ratings but do not contribute to the community by providing their own. In addition, it is difficult for users to rate items if they do not spend enough time evaluating them (especially for resources that need reading such as articles, news, and web pages).

On the other hand, implicit ratings are inferred by monitoring user actions with items to define users' behavioral (or usage) patterns, for example, time spent on a web page, bookmarking, downloading, annotating, printing, or highlighting or sending text to a friend (Beel & Hentschel, 2009; Torres, McNee, Abel, Konstan, & Riedl, 2004). Since implicit ratings are based on the usage of items, sometimes the term "usage-based" ratings is used. The implicit ratings can be used as a solution to data sparsity problems and can be aggregated offline at any time. However, the implicit rating is not accurate and cannot replace the explicit rating (Adomavicius & Tuzhilin, 2005a). The popular example of usage-based ratings is the Google PageRank to evaluate the impact of a web page based on how many other pages refer to it.

- **Overall ratings versus multi-criteria ratings**

Most CF algorithms use one single rating value that represents the average of all users' ratings and is called a *global* or *overall rating*. The overall rating is easier to calculate, but it is not flexible enough to provide adequate details about the quality of the rated item/service. The inflexibility of global rating produces biased recommendations because two users may give the same global rating from two different perspectives [5]. For example, two users may rate a restaurant the same, but the first user rates the food quality while the other rates the variety of food or the restaurant's environment (e.g., furniture, lighting). For this reason, some RSs are based on CF algorithms or hybrid approaches that use multi-criteria ratings based on two or more perspectives (dimensions). For example, in the Papyres system, paper recommendations are based on ten different evaluation criteria that are entered by users (e.g., originality, readability, organization, literature review). The EntreeC restaurant recommender [16] uses cuisine, price, quality, and atmosphere as attributes to evaluate restaurants. The multi-criteria RSs that are based on many ratings are also known as multidimensional or context-aware RSs. Multi-criteria rating RSs will be discussed in section 2.2.

The main shortcoming of CF can be summarized in the following points:

1. *Data sparsity problem:* In collaborative filtering, a user-item matrix is constructed where the columns and rows of this matrix represent users and items. The cells of the matrix contain the ratings that given by users to items. The matrix is called sparse if most of the cells are empty, which means it has very few ratings. The data sparsity problem can occur in many situations. One of the main situations is the case of cold-start problem which caused by new users or new items entering the system. New users are unlikely given good recommendations because they did not rate items yet, and new item cannot be recommended to users until some users rate it. The cold-start problem can be found in some references as new user or new item problem (Adomavicius & Tuzhilin, 2005) [24].
2. *Scalability problem:* with the increase number of users and items which may reaches millions, it is hard to obtain recommendation for all of them and satisfy users' needs for immediate online recommendations.

3. *Gray Sheep problem:* occurs to users who are not consistent in their opinions with any group of users, so that the recommender is unable to classify those users and provide recommendations for them [25].
4. *Shilling Attacks:* occurs when users give unfair high ratings or unfair low ratings intentionally which are not her own actual ratings [26]. This kind of behaviour most often happen in e-commerce applications to give some sellers more good ratings or bad ratings aiming to affect other sellers. Another example is for author cite her own research papers to get more citation accounts. Lam and Riedl [27] found that user-based CF are more affected by shilling attacks than item-based CF, and they suggest that new methods should be used to detect and deal with shilling attacks in RSs.

Breese, Heckerman, and Kadie (1998) identified two main classes of CF: memory-based (continuously comparing all user or item ratings data to produce recommendations) and model-based (a model is learned from the history of item ratings that are used to predict a user's ratings of similar items in the future).

2.1.2.1 Memory-based CF

There are three main types of *memory-based CF*: user-based CF, item-based CF, and demographic CF.

1. *User-based CF:* The recommendation is based on the assumption that the user may have interests and tastes similar to the users who rate the same items similarly. By using the rating activity to find similarities between users, the RS will recommend unseen items to the active user (the user for whom the RS will recommend items) [29].
2. *Item-based CF:* Item-based CF correlates items instead of users, so items that have similar ratings are probably similar [30]. The RS will recommend unseen items to the user if he or she has rated items previously similarly to the unseen ones.
3. *Demographic CF:* In this type of CF, the users are correlated using their attribute information such as age, location, gender, and occupation. Items then could be recommended according to the user's own information [31]. For example, makeup or fashion could be recommended to women.

2.1.2.2 Model-based CF

In *model-based CF*, a model is learned from the history of items' ratings that is used to predict the users' future ratings of similar items. Examples of this CF can be found in Billsus and Pazzani (1998); Breese et al. (1998); Goldberg, Roeder, Gupta, and Perkins (2001); and Hofmann (2003). Building the model is done by applying different machine-learning algorithms such as clustering, Bayesian networks, and rule-based approaches. Breese et al. (1998) proposed two probabilistic approaches in which the probability of the user giving a certain rating to an item is calculated based on a model learned from the previous user's ratings. These two probabilistic approaches are cluster model and Bayesian networks. In cluster model, users who are like-minded are gathered into clusters (classes). The learned model can identify the number of the needed classes. In the Bayesian networks model, each item is represented as a single node in the network while all ratings of each item represent the states of the node in the network. In the rule-based approach, association rules between items are identified. For example, in the e-commerce domain, co-purchased items are analyzed to find the rules of purchasing these items together. Such a model is applied in Sarwar, Karypis, Konstan, and Rield (2000). There are also other probabilistic model techniques such as probabilistic latent semantic analysis [33] and probabilistic matrix factorization (Ma, Yang, Lyu, & King, 2008; Rennie & Srebro, 2005). There are also some RSs that combine both model-based CF and memory-based CF [36].

2.1.3 Hybrid Approach

To alleviate the drawbacks' effects of both of the recommender approaches discussed above, *hybrid approaches* have been used. Hybrid approaches combine two or more recommendation approaches to gain better performance and to avoid the shortcomings of using each recommender approach alone. By using the attributes of the items, we avoid the cold-start problem and data sparsity problem since there is enough information that the recommender system can start with. At the same time, the recommendations' quality usually increases with time in CF. The user can also be aware of items that cannot be seen using the CBF (serendipitous items). Netflix[1] is an example of a movie website which runs a hybrid RSs to recommend movies based first on the user's stated preferences. Later, when the system has collected sufficient data about previous movies watched, similarities between the users and other users' watching habits are used to generate recommendations.

[1] http://www.netflix.com/

There are many ways to combine both approaches [16]. Here we just list them:

- *Weighted score:* Applies content-based and CF separately; then the results given are weighted and combined. The weights are then adjusted during the next recommendations. An example of a hybrid system that uses this combination is the P-Tango system [25].
- *Switching:* In this technique, one filtering is used first, and if it does not give a confident recommendation, the other one is used. The DailyLearner system for adaptive news access used switching to merge content-based filtering and CF. It first employed content-based filtering; then CF was used if content-based filtering started giving unconfident results [37].
- *Mixed:* Produces recommendations from different techniques at the same time. The ProfBuilder system [38] is an example of a mixed hybrid approach.
- *Feature augmentation:* The output from applying the first filtering technique produces a list of candidate items; then this list is input into the other technique. For example, the Libra system [39] uses CF of the data produced by content-based filtering for book information from the Amazon website.
- *Feature combination:* Information from collaborative filtering is used as extra features in content-based filtering. For example, in the movie recommender system in Basu, Hirsh, and Cohen (1998), the user rating of each movie is grouped as a single feature; then content-based filtering is applied.
- *Cascade:* The second filtering technique refines the candidates of the first filtering technique. It applies the second filtering technique only to the candidates that have higher priorities to be recommended and leaves the lower priorities untouched. For example, EntreeC is a restaurant recommender system that uses the cascade hybrid to produce the ranked list. It generates a list of restaurants based on a user's stated preferences. Collaborative filtering is then applied [16].
- *Meta-level:* The first filtering technique generates a learned model that is used as an input for the other. Fab system, an adaptive web page recommender, was the first meta-level hybrid [41]. It uses a content-based recommender to build user models based on weighted term vectors. Then CF is applied to identify similar peers based on these user models and generates recommendations based on ratings.

2.2 Multi-Criteria Recommender System (MCRS)

2.2.1 Definition and Background

Most CF algorithms require users to give just one overall (global) rating (e.g., using a 5-star rating system) and then use the averages of all users' ratings to correlate the items and compute "neighborhoods." This approach is straightforward, but it is not flexible enough to provide adequate details about the quality of the rated item or service. The inflexibility of global ratings produces biased recommendations because two users may give the same global rating from two different perspectives [5]. For example, two researchers may rate a research paper the same, but the first researcher's evaluation is based on the paper's readability while the other's -- on the paper's novelty. For this reason, some recommender systems (RSs) that are based on CF algorithms or hybrid approaches use multi-criteria ratings based on two or more perspectives (dimensions, criteria).

Adomavicius, Sankaranarayanan, Sen and Tuzhilin argued that many dimensions affect recommendations other than just the users and items that are considered in most current RSs [42]. The time, place, and existence of other people are some examples of these dimensions. For example, reserving a room at a hotel could be influenced by many factors, such as the price, cleanliness, and staff friendliness. The user may like or dislike the recommendation produced by the RS depending on factors that cannot be considered only using the users–items matrix. To consider many factors, the matrix should have many dimensions. Each dimension considers one of these factors with different ratings instead of one rating for the item [23]. Recently, some online systems have been using multi-criteria rating systems on sites such as eBay, where buyers rate the sellers according to four various dimensions (item as described, communication, shipping time, shipping charges), and Tigerdirect.com, an online electronics shopping website, which allows customers to rate the products on four different criteria: value, features, quality, and performance. However, these multi-criteria rating systems are not used for personalization but to help customers make decisions about buying products when they view the ratings given by others in the different dimensions.

2.2.2 Classification of MCRS

In a recent review on the research area of MCRSs, Adomavicius, Manouselis and Kwon defined the recommendation problem as multi-criteria decision-making (MCDM) [43]. Most of the surveyed studies can be classified into one of three main categories:

1- Systems that exploit attributes that describe the items to find the users' preferences depending on items the user has liked in the past. Various CBF and knowledge-based systems fit into this category. For example, if a user usually watches comedy movies, the system can filter out other movies based on genre and, subsequently, recommend more comedies to the user.
2- Systems that enable users to specify their preferences about the items' content attributes through searching or filtering processes. The system then recommends items that fulfill the search or filter criteria and are similar to the user's preference. The recommendation is then produced by selecting films that have the features that match the user's preferences [44].
3- Systems that allow the users to give a subjective opinion about each item on multiple criteria specified by the system. The system then recommends items not seen by users that reflects their preferences across these multiple criteria, using their ratings and the ratings provided by other users. Unlike the single-rating systems in the previous two categories, systems in this category use multi-criteria ratings, and each criterion is rated by the user. Thus, users get personalized recommendations based on their previous ratings in all considered criteria.

All of the above categories of systems can be seen as MCRSs because all model the user's preferences as multiple attributes of the items' content. However, the MCRSs in the third category are seen as a new promising trend for the next generation of RSs [45]. Multi-criteria ratings of the quality of items can enrich the user's model and provide extra information.

2.2.3 Differences Between the Single-rating Matrix and Multi-criteria Matrix

Conventional CF approaches arrange the data about users and items into a two-dimensional matrix (see Table 2-1), assuming that the different users are represented in the rows of the matrix and the items are represented in the columns of the matrix. Each cell in the matrix represents the specific user's overall rating, which is a single value, of the specific item that are corresponding to the cell. For example, user A's rating for item 2 is 3.

A rating could be binary (1, 0), which could represent purchasing or not purchasing an item, or it could be an integer or real number that represents the quality rating of the item given by the user. Table 2.1 shows the User x Item matrix containing ratings on a scale between 1 and 5. In order to predict the rating of the target user for item 3, we first must identify the most similar users to the target user (i.e. those who have given other items similar ratings to those given by the target user). In this case, User B is most similar, but User A could also be considered somewhat similar. The prediction for item 3 will be produced by combining the ratings given by the similar users (in this case, say B and A, so we can predict that the rating of the Target user for Item 3 would be close to 4).

Table 2-1: User x Item matrix when overall rating is used (adapted from (Adomavicius et al., 2011))

	Item1	Item2	Item3	Item4	Item5
Target user	**2**	**3**	**?**	**5**	**1**
User A	1	3	3	5	2
User B	**2**	**3**	**4**	**5**	**1**
User C	3	3	4	4	1
User D	2	4	1	3	3
.	.	.	.	.	.
.	.	.	.	.	.

In multi-criteria rating systems, the rating for each item is a multi-value. The matrix for multi-criteria ratings could be represented as follows:

$$R: \text{Users x Items} \rightarrow R_1 \text{ x } R_2 \text{ x } R_3 \text{ x } \ldots. R_n$$

where n is the number of criteria considered in rating the items, and $R_{1,}$ R_2 , . . R_n are the ratings of each individual criterion. In some multi-criteria rating systems, the item's overall rating is also considered. When the overall rating is considered, the utility function could be written as:

$$R: \text{Users x Items} \rightarrow R_0 \text{ x } R_1 \text{ x } R_2 \text{ x } R_3 \text{ x } \ldots. R_n$$

Table 2-2 shows what the multi-criteria matrix looks like; each cell in the matrix contains the multiple values of the item's ratings as well as the overall rating (if there is one). If the task is to predict the target user's rating of item 3, as the table shows, user C is the most similar to the target user considering the overall rating (the first number in bold). However, there are differences between the users' ratings when the multi-criteria ratings are considered. This is to say, if two users give similar overall ratings to the same item, they might give different ratings to different criteria. However, considering the overall rating of the items is helpful because ranking items to be recommended to a user will be very complex if there is no global value that judges the rank [43]. Figure 2-1 shows the rating matrix in 3-D space.

Table 2-2: User x Item matrix in multi-criteria rating system (adapted from [43])

	Item1	Item2	Item3	Item4	Item5
Target user	**2,(**1,2,3)	**3**,(2,3,4)	**?**(?,?,?)	**5**,(5,4,4)	**1,**(1,1,4)
User A	**1**,(1,2,1)	**3**,(3,2,2)	**3**,(3,2,1)	**5**,(3,5,4)	**2**,(3,2,1)
User B	**2**,(3,2,1)	**3**,(4,3,2)	**4**,(3,4,2)	**5**,(3,5,3)	**1**,(3,2,1)
User C	**3**(4,2,2)	**3,**(2,3,2)	**4**,(4,3,3)	**4**,(5,1,5)	**1**,(3,1,2)
User D	**2**,(3,2,1)	**4**,(5,4,3)	**1**,(3,2,1)	**3**,(4,1,3)	**3**,(4,3,1)
.	.	.	.	.	.
.	.	.	.	.	.

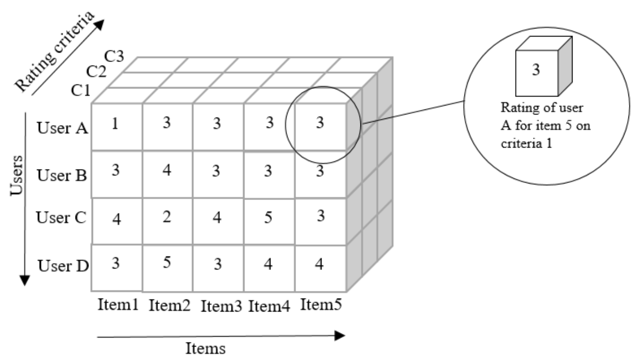


Figure 2-2-1: Multi-criteria rating matrix depicted in 3-D space, user X item X criteria

2.2.4 Approaches of Incorporating Multi-criteria Information in the Recommendation Process

Engaging multiple criteria requires new recommendation techniques that is discussed in this section, in which the recommendation algorithms found in the literature is categorized and examples of some systems that employ those techniques are given.

In order to produce recommendations, two phases are needed:

- *The prediction phase:* in which the prediction of ratings to unknown items is calculated; and
- *The recommendation phase:* in which the decision is made to present the recommended items to the user. For example, the predictions of items are used to rank the items and select the top-N items that increase the user's utility.

The literature shows that incorporating the multi-criteria ratings into the prediction phase can be done according to memory-based and model-based approaches because the conventional CF has the same two approaches. After predicting all unknown ratings and considering overall ratings, the recommendation will be straightforward because it will follow the same approach as if a single rating was used. This is to say, the overall rating is used to select the highly predictable items. However, if the overall rating is not considered, it is a challenge to do the recommendation phase as it is described above[2]. In the following, the approaches for incorporating the multi-criteria ratings in the prediction phase are described.

2.2.4.1 Memory-based Approaches

In the conventional overall rating approach, the prediction is produced using the target user's previous ratings and the ratings of similar users (his or her neighborhood). In multi-criteria CF, Adomavicius et al. [43] found that there are two similarity-based techniques that can be used to consider multi-criteria ratings. The first is to calculate the similarities between users by aggregating traditional similarities from individual ratings of different criteria such as the work done by Adomavicius and Kwon [6] and the work done by Tang and McCalla [46]; the second approach is to use multi-dimensional distance metrics such as Manhattan, Euclidean, or Chebyshev [6].

2.2.4.2 Model-based Approaches

The model-based predictive model does not use the whole dataset in the prediction; rather, it uses a learned model from the observed data in the next predictions for unknown ratings. Many multi-criteria-rating RSs use model-based approaches as discussed below; existing model-based approaches can be classified into:

(a) **Aggregation function:** an aggregation function represents the relationship between the overall rating and the multi-criteria rating, assuming that the overall rating is an aggregate of multi-criteria ratings. The user's overall rating is affected by his or her preference of one or more multi-criteria ratings. For instance, the overall rating of a restaurant with very high rating in the affordable price criterion tends to be very high regardless of the ratings of other criteria. Adomavicius et al. [42] proposed a multidimensional recommendation model (MD model) that calculates the rating of items based on multiple rating values for all considered dimensions.

(b) **Probabilistic modeling:** Some MCRSs adopt a probabilistic modeling approach, which is a machine learning and data mining technique. [47] extended the flexible mixture model

[2] For more information about the approaches used in recommending items without considering the overall ratings, refer to Adomavicius et al. (2011)

(FMM) developed by [48][3] and discovered a correlation between multi-criteria ratings and the overall rating. Experiments using a dataset compiled from Yahoo! Movies showed that when they use very little training data using multi-criteria rating, there is prediction improvement.

(c) **Multi-linear singular value decomposition (MSVD):** SVD techniques have been studied in single-rating RS applications, and they prove their effectiveness in improving recommendations (e.g., Sarwar, Karypis, Konstan, and Riedl, 2000). SVD methods are used as a decomposition method for two-dimensional data (i.e., user and item). In MCRSs, MSVD is also used as a decomposition method to reduce the dimensionality of the multi-criteria ratings. For example, Li, Wang, and Geng (2008) used MSVD in the context of a restaurant RS where users are able to rate each restaurant using ten different criteria (e.g., cuisine, ambience, service). They used the MSVD to reveal the hidden relationships among users, restaurant, and criteria. The information discovered from applying MSVD was then used to locate the user's neighborhood and compute the top-N recommendation. They tested their approach with a dataset that includes 200 users and 30 restaurants. Li et al. found that their approach improves the recommendation accuracy by 5% using the precision measure of top-N in comparison to the conventional single-rating CF.

Even though MCRSs share similarities with context-aware RSs [42] and content-based RS [50], they are different because context-aware RSs and content-based RSs use the objective content attributes (e.g., the length of a movie) or contextual data (e.g., location of theatre or the time of the week), but MCRSs are based on the subjective opinion of the user about each rating criteria.

2.3 Social Network-based Recommender Systems (SNRSs)

Traditional RSs typically use the users' ratings assuming that users are independent and are equal to one another. However, users are influenced by other users (i.e., friends, colleagues) and also have different interests and knowledge as well as different social roles and contexts. Different people have different social relations, which are based on the user's roles of varying kinds and strengths. With the online social network revolution, users have their own connections to people with whom they feel they have the same interests, belong to the same disciplines, or belong to

[3] A method to cluster items and users together simultaneously, so a user or item can be assigned to different clusters at the same time

the same actual community (e.g., company, university, club). Users tend to ask their friends or family members about items, products, or services they may like but cannot decide among because they have similar ratings. For example, if a user likes to watch comedy movies and there are two comedy movies with similarly high ratings, the user may ask his or her friends if they have seen that movie and ask for advice.

Even though the first SNRSs appeared as early as 1997 [51], there was no agreed-upon definition for social recommendation until 2013, when Tang, Hu and Liu gave two definitions, one narrow and one broad [15]. According to the narrow definition, a social recommendation is any recommendation that includes social relations as an extra input to improve the recommendation. TidalTrust [52] and SoRec [53] are examples of SNRSs that follow this definition. According to the broad definition, social recommendations are those made by any recommender who produces recommendations to be used in social media domains, including not only items but also other objects such as people, tags, and communities. Some SNRSs that use the broad definition are Flickr RS [54] to recommend tags and Twittomender [55], which recommends people to follow on Twitter. This study adopts the narrow definition, and all the studies discussed in this section will be ones that follow this definition.

The SNRS uses two matrices, a user–item matrix and a user–user matrix, rather than the single user–item matrix used by the traditional CF. The user–item matrix connects each user to an item (e.g., a research paper) through the user's behavior towards the item (e.g., rating, purchasing, bookmarking). The user–user matrix represents the social relations between users, including *undirected social relations*, such as friendship (e.g., friends on Facebook), co-membership (e.g., membership in the same Facebook group), collaboration (e.g., co-authors of a paper) or colleague relations, and *directed social relations* such as following relations (as on Twitter), watching relations [12], or trust relations [52][56].

Many studies prove that using social information in the recommendation process enhances prediction accuracy [57][58], reduces the effect of the data sparsity and cold start problems [59][53], and increases the user's satisfaction. SNRSs can be classified in various ways such as by the type of social network (explicit or implicit network) or the recommendation approach (memory-based or model-based). Because one of main topics in this research is exploiting the implicit social relations in the recommendation, we classify the related work in SNRSs into explicit and implicit SNRSs. Table 2-3 shows a summary of the related work in SNRS area.

Table 2-3: Comparison between different social recommender systems

Reference	Recommendation approach (input)	Recommending what?	Objective	Dataset used
Explicit social network-based recommender systems **Taste domain**				
Twittomenter [55]	CBF,CF (search terms, social connections (followers, followees), tweet history)	People to follow on Twitter	Compare recommendation using search terms (CBF) with recommendation using the user's history of tweets and social connections	Data gathered from Twitter
[60]	Social CF (users favorite items, friends who are manually selected, user's interaction log with friends)	TV/Movie Items	Compare three different neighborhood formation techniques based on social relations	Facebook
[61]	CF (User behavior (buying/browsing history), the user's friends list)	Books, movies	-Compare recommendations by online RSs and Friends (recommendation accuracy) - Compare interface of six online RS (user's satisfaction)	Books (Amazon.com, RatingZone, Sleeper) and Movies (Amazon.com, MovieCritic, Reel.com) 19 people participated in the experiments
[62]	Social CF (items' ratings, user's social connections)	Local Clubs in Munich	Compare conventional CF with social CF	Lokalisten (German website)
[59]	Probabilistic model-based CF for SNRS (reviews of restaurants, user's friends list and their reviews)	Restaurants	Compare their proposed algorithm using direct/indirect relationships with other algorithms (friend average, weighted friends, Naïve Bayes, and conventional CF)	Yelp.com (restaurant)
[63]	Combination of CF and social relations	Movies	Compare recommendations from user's known friends with recommendations using profile or rating similarity	60 participants hired to simulate online RS
[64]	Random Walk graph model, weighted neighborhood similarity matrix (users' social connections, users' music play counts)	Music	Compare fusing CF with the two proposed methods of social relations	Last.fm
[58]	CF, social CF, hybrid (social CF then CF), or amplifying data of social network in nearest neighbors (users' items ratings and their social connections)	Skin items	Compare four recommendation approaches	Cyworld, a social networking Website

Reference	Recommendation approach (input)	Recommending what?	Objective	Dataset used
[65]	Random Walk with Restart (RWR) algorithm applied for a graph that represent user-user connections, tags and music tracks (Users' music play counts, their tags and their social connections)	Music	Compare the Random Walk with Restarts model and a user-based CF method using the Pearson Correlation similarity.	Last.fm
[57]	Social CF (users' social connections, listened and tagged tracks)	Music	compare the performance of CBF, CF recommenders with their proposed social recommender	Last.fm
[66]	Content matching, content-plus-link, friend-of-friend and SONAR algorithms	People	Evaluate four recommendation algorithms to test their effectiveness on recommending people, and also test if these algorithms increase the number of friends	Beehive system for IBM
Explicit social network-based recommender systems **<u>Academic domain</u>**				
PubRec [67]	Hybrid using the content similarities between papers, popularity of the papers among the user's connections and the numeric rating of the papers given by the user's social connections (users' social connections, users' bookmarks, user's tags and frequencies)	Research papers	Compare proposed PubRec RS with two baseline recommender systems: Social Recommender (SR) which is CF and Tag Vector Similarity (TVS) which is CBF	CiteULike
PReF [68]	Hybrid : CBF (word-correlation factors), then CF using social relations (ratings of user's social connections, user's favorite books, tags)	Books	Compare the recommendation produced by PReF with recommendation from Amazon and LibraryThing	(Amazon and LibraryThing)
PReSA [69]	Hybrid: content similarities of papers considering tags, title, and abstracts; and the popularity of the papers among the user's connections (users' social connections, users' publications)	Research papers	Test the effectiveness of PReSA by comparing different alternative implementations of PReSA, and test the efficiency of PReSA by comparing it to other recommenders (SR, TVS, Cos, Fusion, and PubRec)	CiteULike
[12]	Fusing of watching relations recommendations with CF (users' bookmarks and their watching users)	Research papers	Compare the accuracy between conventional CF, social, and the hybrid approaches in terms of precision and recall	CiteULike
[70]	Fusing of group memberships	Research papers	Compare the accuracy between	CiteULike

Reference	Recommendation approach (input)	Recommending what?	Objective	Dataset used
	recommendation with CF (users' bookmarks and their group membership)		conventional CF, group-member-based, group-based, and the hybrid approaches in terms of precision and recall. Then conducted user-study with 8 users	
[71]	Community vote-based, CF, hybrid such as content-boosted social network (CBSN), CBSN with social features, content-boosted CF (users' publications, talks bookmarks,	Conference talks	Explore different algorithms in conference talks domain for their recommendation accuracy and alleviating the cold start problem.	Conference navigator dataset for two conferences: ASIS&T 2010, iConference 2011, Scopus to retrieve the users' publications
Implicit social network-based recommender systems **<u>Trust-based social networks</u>**				
[72]	Graph-based method. Eliminating the graph cycles then apply graph walk over the social network (users' FOAF (Friend Of A Friend, ski resort routes)	Ski resorts	Implement a new trust metric MoleTrust	Moleskiing.it
[73]	Trust-based recommendation using the propagation of trust between users (item ratings and users' trust ratings)	General items	Compare the trust-based recommendation to the CF and study the effect of trust propagation on cold-start users	Epinions
[74]	Combine trust and similarity to produce compound weighting to be used in the prediction or using trust to include only trustworthy profile in the recommendation (profiles of movie ratings)	Movies	Compare different recommendation strategies (profile-level trust/item-level trust, trust-based weighting/trust-based filtering) to test the gained benefit and mean error rate	MovieLens dataset
[75]	Uses social network trust inference algorithm called TidalTrust (item ratings and users' trust ratings)	Movies	Compare the trust-based prediction to the conventional CF	FilmTrust
SoRec [53]	Latent factor analysis using probabilistic matrix factorization (item ratings and users' trust ratings)	General items	Compare the algorithm with other matrix factorization methods	Epinions
[76]	Random walk algorithm, combination of content-based and trust-based (item ratings and users' trust ratings)	General items	Compare the proposed algorithm (TrustWalk) with other trust-based methods (e.g. TidalTrust, MoleTrust), with user-based, item-based CF	Epinions
Implicit social network-based recommender systems **<u>Non_Trust-based social networks</u>**				

Reference	Recommendation approach (input)	Recommending what?	Objective	Dataset used
[77]	Behavioral Network Collaborative Filtering (BNCF) (users' navigational activities)	General items	Compared with CF, and with the approach that uses only direct social relations	Usage dataset from intranet of Credit Agricole Banking Group
[78]	Fusion of trust values and item ratings (item ratings and users trust ratings)	General items	Compare the proposed credibility-based approach with conventional CF and trust-based approach by Golbeck (reference)	Epinions
[79]	Hybrid approach: integration of CB characteristics into a social network-based CF system (keywords of the conference topics, program committee members data)	Experts	Investigate the performance accuracy of the recommendation system with and without the social network component	Crawled dataset for 315 program committee members of the 16th ACM SIGKDD conference
[80]	Construct implicit social networks based on the co-occurrence keywords and names appears in the user's publications (user's publications)	Researchers and research papers	Calculate the precision is as the percentage overlap between recommendations and the target user's existing friends-list extracted from SCHOLAT	SCHOLAT dataset
[81]	Social CF using four proposed implicit social relations (user's item ratings, user's review comments)	Movies	Compare three recommendation methods	Amazon
Implicit and Explicit social network				
[82]	Use of familiarity, similarity networks and combination of both to recommend items related to people within the user's network (users' bookmarks, comments, tags, and their social connections)	Bookmarks of Internet and Intranet	User study to compare the three social networks with/without explanation	Lotus Connections dataset
(WSNRS) [83]	Fusion of trust value with rating value (Users' social connections and the type of the connection, number resources published by the user, comments, ratings, clicks and other data represent user's interest)	Recently published resources	Case study only provided	Dataset from Intelepciune.ro website
[84]	Recommend people based on the strength of the relations in each of nine interaction layers: users' social connections, users' interactions with photos (comment, tag, add to favorites)	Users	Compare the recommendation using equal weights for proposed layers with the one that uses personal weights adjusted for each user	Flickr

2.3.1 Explicit SNRSs

The relations in a network are called explicit when they are initiated by one user or with the agreement of both users involved in the social relation. Examples of explicit relations are friends on Facebook, or Twitter/Instagram followers. Many recommender systems are based on explicit social networks, and most are based on friendship relations. The research on exploiting explicit social relations in RSs is applied mainly in two domains: taste and academic. We first discuss the studies in the taste domain, such as recommending music or movies, then we discuss the work that has been done in the academic domain such as recommending books, research papers, or authors with whom to follow or collaborate. The first part of Table 2-3 summarizes the SNRSs that exploit explicit social relations.

He and Chu developed a probabilistic model to produce personalized recommendations by utilizing information found on a social network, including the user's item preference, general acceptance of the item, and the opinions of the user's friends [59]. By investigating users' friends' reviews, the authors found that friends tend to visit the same items and have similar ratings to the target user. They argued that the integration of social information can increase the performance of traditional RSs for three reasons. First, it increases prediction accuracy because modeling users can be more precise due to having a better understanding of the users' behavior and ratings. Second, it reduces data sparsity because users who are connected most likely have something in common, so there is no need to find additional similar users. Finally, it reduces the effect of the cold start problem. The user's friends' ratings can be used even if the user has no ratings yet. He and Chu showed that using social networks improved prediction accuracy by 17.8% in comparison to conventional CFs.

Twittomender is another SNRS that recommends people to follow on Twitter [55]. Twittomender has two main recommendation possibilities: using the search terms to recommend people with similar interests or using the active user's tweets and social connections (i.e., who he or she follows and is following). In [66], four recommendation algorithms to recommend people were evaluated: one is content matching which is pure CBF, Content-plus-Link which is a combination of CBF and recommendation using the social links, Friend-of-Friend algorithm that exploits the relations between directly connected friends to recommend new people using the propagation of the social relations, and the last is the SONAR algorithm which is based on SONAR system that aggregates social relationship information from different public data sources. Researchers found that all algorithms were effective in increasing the number of friends, and while the recommendations based on social relations were good to find known people to the target users, the recommendations using the content matching were stronger in finding new friends.

Bourke, McCarthy, and Smyth examined three neighborhood formation techniques based on users' social relations to amplify friends' opinions about TV and movie items and compared the results obtained from the proposed techniques to each other and to the results from the conventional CF [60]. The first two techniques of selecting friends are (a) enabling users to manually select the friends and (b) selecting friends automatically from the list of friends based on the frequency of the interactions. The third method is similar to the second method but considers the Jaccard coefficient similarity between the target user and his or her friends.

Bonhard, Sasse, and Harries conducted an experiment with 60 participants to explore film recommendations from participants' real-life friends and comparing them with profile and rating similarities [63]. The authors found that participants prefer their friends as recommenders, and this preference increases if they have higher profile and rating similarities with their friends. They also found that ratings from friends with whom the user shares more profile and ratings similarities are more trusted. They also proposed an interface that can be used to highlight and enlarge the keywords that are the most common between the user and his or her friends. However, the effect of the user interface was not tested.

Konstas, Stathopoulos, and Jose applied the random walk by restart (RWR) algorithm and compared it to a user-based collaborative filtering method using the Pearson correlation similarity [65]. The authors collected the user's explicitly expressed bonds of friendship and their tags from Last.fm, a music-centred social network. After conducting a series of experiments, they found that the RWR benefited from the incorporation of the social information and outperformed the standard CF method in precision and recall. However, their data collection method is questionable because they collected the top 50 fans of the top 50 musicians; applying the same method to randomly chosen users might not lead to the same results.

In [64], researchers tested the effect of two explicit social networks, membership and friendship, when fused with conventional CF recommendation methods. Their study was also conducted using information from Last.fm. The authors compared two kinds of fusion CF with social relations: one with a Random Walk graph model and the other via a weighted neighborhood similarity matrix. The study showed a significant improvement in recommendation accuracy (i.e., precision, recall, and F-measure), which increased by up to 8% when the graph model was used.

Another study done by Bellogin et al. used different approaches to test the performance of different recommendation approaches not only using performance measures, such as precision, recall, normalized discounted cumulative gain (NDCG), but also nonperformance measures, such as coverage, diversity, and novelty of the recommendations [57]. The results showed that, in a dataset compiled from Last.fm, tagging and explicit social network information produced effective and heterogeneous music recommendations. [61] compared six online RSs (three books and three movie) with recommendations from friends, and they found that friends make superior recommendations than online RSs. They also found that the user’s satisfaction was higher with friends’ recommendations.

Using a dataset compiled from Lokalisten, a Munich-based German-language virtual community, [62] found that recommendations of local clubs using social CF, which considered the user’s friends to form the user’s neighborhood, outperformed the recommendations produced using the conventional CF neighborhood. However, the results are significant only in specific cases, such as when there are very sparse ratings or novel predictions are needed. They also found that friends’ ratings are correlated in comparison with non-friends.

In [58] four algorithms were compared: nearest neighborhood CF, social CF, a combination of nearest neighborhood CF and social CF, and nearest neighborhood CF with an amplification of data from social friends. The authors collected each user’s preference ratings and friendship relations from the South Korean social network Cyworld, and they found that combining data from nearest neighborhood CF and social CF perform the best in terms of MAE.

SNRSs have also been tested in the academic domain, including the recommendation of expert or research papers. For instance, PReF is a book SNRS that uses a hybrid recommendation approach [68]. Pera and Ng used data compiled from LibraryThing.com, social website; first, they applied a CBF based on word-correlation factors to find books with similar content. They then applied a CF based on the user’s friends’ ratings. They found that their approach outperformed the CF recommendations provided by Amazon.com and the CBF provided by LibraryThing.com in terms of precision and ranking. Pera and Ng also proposed PubRec [67], an RS that suggests the most related papers for a particular paper from a list of papers available in the libraries of users who are socially connected to the target user. PReSA [69] is another SNRS that takes advantage of the available data on social bookmarking websites (e.g., CiteULike), such as bookmarked papers, metadata of papers, and users' connections, to recommend similar papers from the users' connections' libraries that are popular among the users' social connections. Both PubRec and PReSA consider the explicit relationships among users in the recommendation process.

Lee and Brusilovsky [12] [11] [71]have studied three explicit social networks to find the extent of interest similarities between users involved in those networks. The three networks are a watching network [12], a group membership network [11], and collaboration networks [71]. Watching relations are unilateral relations initiated by one user to watch (or follow) the other user's library updates. Lee and Brusilovsky [12] have studied the recommendations produced by watching networks and compared them to the traditional CF. Their results showed that the watching network cannot compete with CF. Group membership networks connect users who are members of the same group. Even though users agree about the relevance of the group topic to their interests, the similarities between users' libraries in these networks is insignificant: 0.29% of item similarity, 0.83% of metadata similarity, and 0.86% of macro-tag similarity [11]. The collaboration network connects two users if they coauthor a paper [71]. The similarity between two users connected using a co-authorship network is comparable to that of social connections that need agreement between the two partners involved. However, the results showed that the collaborators' similarities were lower than the similarities between socially connected users.

2.3.2 Implicit SNRSs

Implicit social relations are inferred from user behavior such as browsing, bookmarking items, tagging, and commenting on items. These relations are constructed by machine calculations without user intervention, and so, users in implicit social networks may not be aware that they are socially connected; in this way implicit social networks are similar to the neighbourhoods in CF. Implicit social relations are beneficial finding better recommended items and alleviating some of the recommendation problems, since datasets that contain explicit social connections are rarely (if ever) found. In fact researchers are forced to either crawl through the data themselves or build implicit social networks based on the available data.

The literature contains some works that use trust to infer relations between users. All SNRSs use trust to some degree. Some use the assumption that the user is usually influenced by the behavior of other people in his or her social network, while other SNRSs calculate trust values between two users who are connected and adjust these values each time the users interact. In this section, the work done on implicit social networks is classified as trust-based and non-trust-based. See table 2.3 for the summary of the work that is done in both.

2.3.2.1 Trust-based

The notions of trust and reputation are often used interchangeably. However, they often differ in the way they are evaluated. According to [85], trust is a person's belief in another person's capabilities, and reputation represents an aggregate belief about another person's capabilities based on the opinions (or trust) of others. Trust and reputation share some characteristics. First, trust and reputation are context-specific, which means that, for example, a person may trust another person to be a good teacher but not a good football player. Second, trust and reputation are transitive. An example would be a case in which there are three users A, B, and C, where users A and B are friends, and users B and C are also friends, but users A and C are not connected directly. Using the transitive property, we can conclude that user A would trust user C because user A trusts user B and user B trusts user C, so, subsequently, A also trusts B's friends by reference. Third, trust and reputation may have different aspects. For example, a user may evaluate a teacher based on the quality of the teacher's explanations, patience, and willingness to help.

Trust values can be acquired explicitly or implicitly. Explicit trust values can be acquired through the evaluation of the trustee's explicit response toward the services provided by the trustor, such as the RSs in [75] and [73]. Trust values can also be acquired implicitly by using the user's behavior or other information that relates users to one other, such as ratings [59], or by using user or item profiles [74]. In addition, the user's trust value could be inferred by monitoring other users' behavior toward the trustee. For example, the number of followers indicates how much the person is trusted in a specific context.

It is evident that users prefer to ask for and find recommendations from people who they trust rather than using traditional RSs [61]. In social networks, users accept recommendations for items that are highly recommended by people who they trust [86]. For that reason, many scholars have tested the effect of constructing trust-aware RSs. The FilmTrust social website system proposed by Golbeck [75], for instance, reviews and rates others' movie tastes and recommends movies using the trust developed between users based on similar movie ratings.

Trust is mainly used to reduce the weaknesses of traditional RSs. Many proposed trust-aware recommendation algorithms have shown that the performance of traditional RSs can be improved by including the trust relationships between users [53]. Avesani, Massa, and Tiella [72] proposed a trust-aware RS that uses local trust metrics to personalize recommendations for secure skiing routes (Moleskiing RS). Personalization only shows information from users the target user trusts, which is not the case in systems based on global trust values, such as Google PageRank. The trust factor in Moleskiing is used to alleviate the data sparsity problem by utilizing trust propagation to infer trust values for unknown users. This approach can generate predictions for 66% of users while the conventional CF could only predict for 14% of the users. [73] proposed a trust-graph-based RS that uses trust relationships between users in addition to the similarity of ratings to reduce the data sparseness that affects users, especially new ones. The results of the experiments performed on the Epinions dataset showed that trust-aware RSs outperform CF in terms of recommendation accuracy. While this study uses explicit trust values, the study done by O'Donovan and Smyth used implicit trust values that were inferred from user ratings [74]. The trust values consisted of profile trust and item trust. While Massa and Avesani [73] aimed to overcome technical problems with CFs (i.e. data sparsity), O'Donovan and Smyth [74] focused on increasing the accuracy of recommendation predictions. To find a solution that fulfills both objectives—reducing data sparseness and increasing prediction accuracy—Ma et al. [53] proposed the SoRec system. SoRec integrates the user's social network graph with the user-item rating matrix using latent factor analysis and probabilistic matrix factorization. The social network used is a trust-based social network where users can classify other users as "trusted" or "blocked." The study's main objective is to fill the empty cells in the user-item matrix by finding the probability of the user-rating prediction in order to decrease the matrix's sparseness. One of the main advantages of Ma et al.'s approach is that they dealt with the confidence of the trust values. In other words, if the user trusts many users, his or her confidence of trust is decreased, but when the user is trusted by many users, his or her confidence of trust is increased. Ma et al. tested this method using a dataset compiled from Epinions; results showed that the SoRec outperformed other matrix-factorization-based CF techniques and it can produce recommendations for new users or users with few if any ratings.

2.3.2.2 Non-trust-based

Work has also been done in the area of exploiting the implicit social relations in the recommendation rather than the work the uses trust. In [78], a study was proposed a “follow-the-leader” model that classified users into leaders and followers. Leaders are users who have ratings that are minimally different from the average. The authors’ algorithm collected trust values from the leaders’ direct and indirect followers and recommended items based on the credibility of leaders in the system. They replaced the similarity weight with the credibility weight, and if the user was new and did not have enough ratings, the credibility weight was used instead of the trust weight to calculate the predicted ratings of the unknown items. The authors’ experiment with the dataset compiled from Epinions showed that their approach could predict ratings for new users and also increase the coverage. On the other hand, it could decrease personalization and allow users to game the system to become leaders.

Davoodi, Afsharchi and Kianmehr [79] developed an SNRS to recommend experts who match the topics in which the active users need help. Their method identified representative expert communities by examining similarities in experience, background, knowledge level, and personal preferences. A clustering algorithm is used to build a social network of experts. Scholarmender [80] is an RS that recommends researchers and research papers. The algorithm first collected the names of the papers’ authors and the papers’ keywords. A social network of researchers was then constructed by finding the similarities between two researchers; the more the scholars coauthored papers and had overlapping keywords, the more similar they were considered to be. Chen et al. conducted their experiment using a dataset compiled from SCHOLAT, a Chinese social network website for the use of researchers. However, they had low precision for recommendation accuracy in the lists of 10 people and 25 papers that were recommended for each user. They justified their findings by indicating that users had an average of only 2.5 social relations in the proposed implicit social network.

In Esslimani et al. [77], behavioral network collaborative filtering (BNCF) was presented when implicit relationships between users were inferred through their navigational behaviors within the network. User navigation patterns were recorded and correlated. The items’ ratings were inferred by using two features: the frequency of visiting a resource and the duration of a visit to a resource. The accuracy of the behavioral approach outperformed both CF and the approach that used only direct links between users, as in [87], because the behavioral approach uses the social network's transitivity to construct new links between users.

In [81], four methods to extract implicit social links were proposed. Two methods are based on users' ratings—common rating and Pearson moment correlation—and two are based on topic similarity—topic similarity and fine-grained topic similarity. The authors compared the results of these methods using three different social recommendations methods. The experiments using a dataset compiled from Amazon showed that fine-grained topic similarity performed the best in comparison to conventional CF and social MF, which uses matrix factorization techniques to allow for the propagation of interest through social relations.

TrustWalker [76] recommends items using a combination of content-based and trust-based recommendations. The authors proposed a random-walk algorithm that attempts to find the trade-off between precision and coverage. Using trust propagation increases the coverage but reduces the precision, because the further one goes through the network, the more one could find less trustworthy raters. For this reason, TrustWalker uses a content-based method to find raters for similar items from trustworthy users instead of finding raters for the exact items but who are less trustworthy.

2.3.3 Considering Implicit and Explicit Social Relations

Some studies compared the recommendations produced by explicit social networks with those produced by implicit social networks. For example, [82] compared a familiarity social network with a similarity social network. A familiarity social network connects users through explicit social relations wherein the users know each other. A similarity social network is based on implicit social relations among users based on their behaviors, such as using the same tag or bookmarking the same webpage. They also compared these two social networks with a network that was created with people who are familiar and similar to other users (i.e., a combination of both previous social networks). They showed that the recommendation from the familiarity social network outperformed the recommendation of the similarity social network. They explained that they got these results because the recommendations are explained to the users, who can see the picture of the contact who sent the recommendation. However, they might get this result because all users belong to the same community, which, in this case, was the IBM Corporation, and users all have similar interests.

A wise social network recommender system (WSNRS) was proposed by Mican et al., which considered explicit and implicit social relations (e.g., implicit relations based on number of clicks to see other user's profile) [83]. First, the algorithm considered the user's connections which are made up of users who have an explicit social relations with the target user. It then considered the interactions between the target user and other users as well as the interactions between the target user and the webpages to calculate a trust value. If the trust value was above average, the target user is an implicit follower of the other user. The recently published resources and the favorably rated resources from the target user's connections are then recommended to the target user. Mican et al. explained and demonstrated this using a case study that was neither evaluated by any evaluation measures nor compared with any baseline recommendation methods. Thus, we cannot conclude the effectiveness of the proposed method.

Kazienko, Musial, and Kajdanowicz defined indirect relationships between users through their interactions with objects [84]. The authors proposed a personalized RS that connects two users directly through the contact lists of Flickr accounts. Users can also be connected indirectly through another user's contact list, through membership in the same group, or through the interaction with photos (e.g., comment, tag or add to favorites). All of these interactions are recorded, and the user of the RS assigns weights to these interactions according to their importance. These weights are also adjusted through user feedback. All the collected data about the social relations are then associated with the relation strengths and used to recommend a ranked list of users to the target user.

2.3.4 Summary

This section discusses the related work of SNRSs. As shown in Table 2.3, most SNRSs are applied in the taste domain such as recommending movies, music, or restaurants. In the taste domain, the users' interest play the most important role. However, it is difficult to use the recommendation algorithms for the taste domain to recommend research papers because there are many factors other than the user's interest that affect one's choice to read a research paper. The user's knowledge level, goals, and context also play important roles in choosing research papers. In addition, in taste domains, the number of ratings for each item is larger than the number of ratings received per research paper. Furthermore, there is a lack of explicit ratings in the domain of research paper recommendation. Most of the research done in this area is based on citation networks and implicit feedback about the papers, such as co-tagging, co-downloading, or co-bookmarking (please refer to section 2.4 for the related work about recommending research papers). For these reasons, our contribution to this research is two-fold. First, we test the users' interest similarities in different implicit social networks that are proposed and constructed, which use data from social bookmarking systems and are used as sources for paper recommending algorithms. Second, we adopt a multi-criteria rating system to build a research paper RS that exploits the multi-criteria ratings and data from implicit social networks.

As shown in Table 2-3, 19.35% of the studies surveyed exploit explicit social relations when recommending research papers or other relevant items such as conference talks or books. However, these six studies are done by only two research groups. However, in these studies, the coverage of their algorithms is low, which means that, in most cases, the algorithms cannot produce recommendations. In other words, people who are not involved in any explicit social relations with other users cannot get recommendations from these systems. For example, when a collaboration network is used, only 25.3% of the user population can get recommendations while the coverage in a conventional CF is 70.7%. The case is worse for the watching network, in which 96.6% of users do not have watching relations and, consequently, cannot receive recommendations, and 91.5% of users do not have group relations [10]. In the studies done by Pera and Ng [67][68][69], the explicit social connections are based on the invitation used. The recommendation that uses this kind of a social connection has also lower coverage than CF. In their studies, Pera and Ng used datasets compiled from CiteULike; however, authors did not take coverage (the number of users who receive recommendation) into their account, and, as a result, we do not know how many users in their dataset do not have any social connections. However, when we collected a larger dataset from CiteULike, we found that only 18% of the users in the dataset have explicit social connections with an average of 0.31 connections per user (please see chapter 4).

Only one study exploited implicit social relations in recommending research papers [80]. However, their approach was based on the user's own publications, which means that the algorithm cannot recommend items for users who do not have publications. The other study related to our research is done by Guy et al. [82], which compared using explicit and implicit social networks when recommending bookmarks. However, their results are questionable because the study was done in a closed community where users are familiar with each other.

Thus, there is a lack of research done on SNRSs in recommending research papers in general. Our contribution to this research has the following features. First, no publication is required for the user in order for research papers to be recommended. The proposed implicit social networks deal with all users equally so that users with or without publications have the same chance to be part of a social network. Second, no data input is required from the user; the publicly available data is used based on data the user provides to the social bookmarking tool during their regular activity with the system. Third, users are not required to have explicit social connections to be part of the proposed social networks.

2.4 Recommender Systems for Researchers

In the previous sections of this chapter, the work that has been done in the MCRS and SNRSs were explained. In this section, we review the area of research paper RSs is provided and we categorize them based on the recommendation approaches that used. Then, we provide a summary to show if research paper RSs used some of the aforementioned areas (MCRS and SNRS).

2.4.1 Recommending Research Papers using Content-based Filtering

Content-based filtering (CBF) is based on information retrieval techniques comparing a paper's features (e.g., title, abstract, keywords, publication year) with the researchers' features (e.g., interests or previous search queries) in order to find matches. For this reason, content-based approaches are also called *feature-based*. Methods for CBF, based on text analysis, are widely used since each paper can be represented as a bag of words, calculating the weight of each significant word in the paper to provide a measure of its importance. The assumption is that the more frequently the word appears in the paper, the more important and representative it is for the paper. Although text-based methods for CBF are widely used, some criticisms have been raised. For example, a problem may arise when using different synonyms to refer to the same term in two papers. Also, the same word could be interpreted differently according to the context it appears in.

Sometimes, full-text analysis is deployed to alleviate these problems [21]. However, it can be done only if the full text of the paper is available, which may not be the case due to copyright issues. Another problem with using the full text is that the processing takes a long time and requires extensive computation. Other CBF methods use only the paper's meta-data such as title, abstract, or keywords or a combination of them. In the algorithm proposed in [88], for example, the text of the abstract was used to find more papers that addressed the same problem or papers that used a similar solution. Text-based CBF research paper recommenders have been used in a wide range of applications. For example, He et al. [89] recommended references for authors who are in the process of writing a paper, enabling them to find the most suitable reference for a particular statement or specific location in the text. In this case, the text around the reference placeholder is processed, and papers with similar content are recommended as references in that location. Basu et al. addressed the problem of assigning conference papers to reviewers by developing a CBF recommender that can run many different algorithms using different data from the papers (i.e., title, abstract, keywords) separately or by combining some or all of them [1]. The authors found that using a combination of the papers' data is more effective and can recommend better matches of papers to reviewers. Zhang et al. [90] proposed another CBF algorithm that aims to find novel papers that contain new information to the user. After finding relevant papers based on text similarity, a redundancy filter discards any papers that are not novel. Chandrasekaran et al. proposed another CBF approach based on semantic indexing of the paper [91], similar to [92]. They represented the user profile and the papers as trees of concepts based on the ACM topics classification, but they computed the similarity between papers to be recommended using the tree-edit distance algorithm [93] to find the distance between concept vectors instead of using keyword vectors. Their results showed that the concept-based recommendation outperforms the keyword-based one. However, the method requires a preprocessing stage to annotate papers with respect to concepts.

Combining text-based methods with other content-based methods is used to increase RS performance. For example, Google Scholar applies text analysis and citation counts (as the number of citations can be considered to be a function of the content of the paper) to find and rank similar papers to the search query. Another approach [92] combines text analysis and citation graphs to construct different researcher profiles (for junior and senior researchers based on their publications). Then text-based matching can be done among the researchers' profiles (represented as a weighted vector of the keywords appearing in the researchers' publication(s)) and the papers in the citation graph. The same authors [94] used a similar combination of text-based and citation-based approaches to recommend serendipitous papers instead of using random selection. They found that constructing a researcher's profile using the citation graph to find less similar papers produces more effective serendipitous papers than random selection. Although [94] classified their work as using a content-based approach, we classify them as a hybrid approach since they used both the text (CBF) and the citation graph, and citations could be considered as implicit ratings (CF). Other CBF recommendations are produced by observing users' viewing history of papers such as the recommender of the Papits system [95].

Purely content-based features may not be sufficient to generate appropriate recommendations. For example, the recommended paper could be relevant to the interests of the user based on the content features (text, semantic annotations), but how influential the paper is could not be discovered, especially if it is a new paper that has no citations yet. In addition, CBF approaches can overspecialize the recommendations, which means researchers would not have the opportunity to read surprising, highly original papers or papers that cover other perspectives on the topic.

2.4.2 Recommending Papers using Collaborative Filtering

In contrast to CBF, the CF approach in general is helpful if few or no features about the items or users are available but user ratings are. CF for recommending papers could be based on readers' explicit ratings, citation analysis, or usage analysis. Due to the lack of available ratings, citation analysis is the most common approach, followed by usage analysis.

2.4.2.1 Citation-based Methods

Citation analysis (or bibliometrics) is used to define the relationships among papers. The method is based on the assumption that papers that cite other papers are in some sense related and similar. Citation relationships can be represented as a graph where the vertices are the papers and the edges are the citing/reference relationships. The most well-known graph-based measure is the Google PageRank [96], which deals with web links as citations. The rank of a page is recursively calculated as a sum of the ranks of the pages that point to that page.

There are four approaches to analyzing citations shown in Figure 2-2. Figure 2.2.A represents the case if paper **a** cites paper **b**. The "cited by" relationship is represented in Figure 2.2.B where **a** is cited by **b**. The co-citation is the relationship between two papers cited by a third paper. Figure 2.2.C shows that **a** and **b** are co-cited in **c**. The co-reference relationship is a counterpart of co-citation, in which papers are considered related if they share one or more references. Figure 2.2.D shows that **a** and **b** are co-referenced.

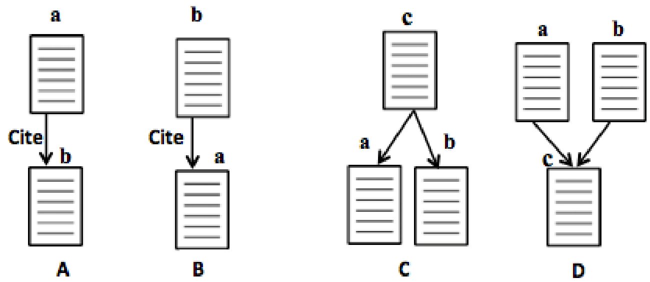


Figure 2-2: Different reference relationships between papers, adapted from [21]

Many CBF algorithms use citation analysis to find relevant papers. For example, CiteseerX,[4] a citation database, was the first to use citation analysis in 1997, specifically co-citation and co-referencing, to find papers similar to a specific input paper. The cites and cited by relationships are used in [92].

The citation analysis is used to build a citation matrix that was first proposed by [2] in place of a matrix with user ratings for papers to accomplish CF of research papers. For example, the entry Cp1,p2 in the citation matrix represents the relationship between two papers (p1,p2). If Cp1,p2 is equal to 1, this means p1 cites p2, i.e., it contains a reference to p2; it would be zero otherwise. While CF RSs typically utilize explicit user ratings of items (e.g., in movie or book RSs), paper RSs use the citation of papers as implicit ratings. This type of implicit rating alleviates the data sparsity and cold start problems inherent in collaborative filtering approaches. In [2], six different CF algorithms were compared within the classical RS GroupLens to find suitable additional references for a target research paper. The results showed that despite few available data from the user, the algorithms produced effective recommendations. In [97], authors have built the paper recommendations based on the paper's citations and the trust of the active users to others users' reviews of the paper. The relationships between papers and users are represented as two layers to connect users to papers through their trust of the reviews.

In addition, analyzing citations enables calculating the citation count, which reflects the paper impact. Thereafter, citation count is used to calculate the h-index of the author, which shows an individual author's influence and long-term productivity (indicative of her authority in a field). However, there are some limitations to using citation analysis in generating recommendations [21]. First, not all papers can be identified correctly as authored by a particular

[4] http://citeseerx.ist.psu.edu/index

individual because authors may share the same last name and initials. Second, not all the papers that appear in the reference list of a paper have been cited because of their relevance. Third, the citation count should be used very carefully for comparing papers, as it depends on the time since the paper has appeared in the area of study and other factors, e.g., nationality of the authors [98]. Fourth, unlike PageRank, the citation count deals with all papers similarly; it considers a self-citation as important as a citation from a paper with a high impact factor.

More recently, PaperTaste, a personalized paper recommender system is proposed in [99]. Authors used six citation paths to collect candidate papers: citation, reference, co-citation, co-reference, co-author, and co-venue. The final candidate papers are the union of the papers generated by the six paths. Then, the candidate papers are ranked using a learned ranking model and the user profile that consider the user different activity in SocialScholar[5], social network for computer science researchers. The algorithm is compared to a PageRank-weighted CF, CBF, conventional CF and fusion of CBF with PageRank. [100] proposed a new way to use the citation network. They applied Greedy Clique expansion algorithm, which inspired by PageRank, to discover communities based on citations. Then, for each paper in each community, the rank of its influence is calculated which means each community has a ranked list of recommendations. Then based on the user's entered keywords, they are mapped to relevant communities and the user will be recommended the list of authoritative papers in that community. [101] proposed an algorithm to predict the target user's interest in newly published papers using the paper's reference list. If the target user has cited or published any of the paper that are cited in the newly published paper, this mean she is interested in this cited paper. Then the prediction for the newly published paper can be calculated by integrating her interests of all of the cited papers in that paper. They compare their model-based CF using SVD (Singular value Decomposition) with two other citation-based methods: the PageRank [102] and belief propagation [103]. The results showed that their algorithm perform better that the other algorithms in terms of accuracy using precision and recall measures.

Another relationship between papers that can be represented as a graph is co-authorship. In this case, authors can be represented as nodes that are connected directly if they co-author one or more papers. This kind of graph is helpful in finding more related papers since authors who

[5] http://soscholar.com/

collaborate in authoring a paper most likely share topics of interest and therefore have authored or have liked more related papers. Graph-based approaches (citation-based, co-authored-based) in essence consist of navigating the graph to find more papers to recommend for other researchers using connections between the nodes in the graph.

2.4.2.2 Usage-based Methods

Usage-based methods refer to the methods inferring the similarity of interests among users by monitoring the actions of researchers, such as co-downloading (e.g., bX[6] RS), co-accessing papers [104], or saving the same papers in personal libraries. The user's actions toward a paper (e.g., viewing, printing, annotating, or sending the paper to other researchers) are monitored to infer the researcher's interest [21]. The CADAL recommender [105] uses the user's access log to books in the CADAL digital library (China America Digital Academic library) to find books that the active user has not read but her neighbors have. The bX is a commercial usage-based CF recommendation service for a large-scale digital library of scholarly papers from many institutions around the world. bX collects users' co-downloads from the log files of the institutions starting from the paper that the user is looking at. Other authors (e.g., Pohl et al. [104]) also use the co-downloaded papers to find more similar papers. Monitoring the actions of researchers is useful to infer their interests without the need of explicit ratings. In addition, it is an effective way to infer the short-term and long-term interests of the researcher. Some of the advantages of usage-based methods are the earlier availability of usage data in contrast to citation data, and they are more accurate for recently published papers that may not be cited yet [104]. However, monitoring researchers' actions continually could raise privacy issues [21].

2.4.2.3 Ratings-based Methods

Many RSs based on "classical" collaborative filtering with user ratings have been proposed. For example, the System for Electronic Recommendation Filtering (SERF) [3] asks users to enter long and informative queries; then it collects users' ratings about the search results, whether the results meet their information needs or not, and then uses these ratings to recommend papers to later users with similar needs. The evaluation results show that users' feedback about recommendations increases both their effectiveness and efficiency. In [106], explicit rating was used to compare user-based CF (that generates neighborhoods of users similar to the active user)

[6] http://www.exlibrisgroup.com/?catid={7098DEDA-0C18-41C0-B9E0-FD36B705EE5D}

and two versions of item-based CF (that generates neighborhoods of papers similar to an active paper). The results showed that user-based CF achieves the best performance; however, two years is needed to overcome the cold start problem. Also, Tang and McCalla proposed a multidimensional recommender system that collects different users' ratings for each paper based on many rating dimensions: value added by reading the paper, the paper difficulty level, and whether the user was going to recommend this paper to other peers or not [46]. Parra-Santander and Brusilovsky also used explicit user's ratings; they made two enhancements of user-based CF for social tagging systems and evaluated them by conducting an experiment to compare these enhancements to the traditional user-based CF [107]. One enhancement was done by using a tag-matching approach called BM25 (the best known keyword-matching technique in the IR field) to match collaborative tags, whereas the other enhancement was related to the prediction stage by using neighbor-weighted CF (NwCF), which takes into account the number of raters in the algorithm to filter out the papers that have been rated by one or two users. Combining both enhancements gives the best precision results.

2.4.3 Recommending Papers using Hybrid Approach

To overcome the drawbacks of CBF and CF, hybrid approaches have been used to recommend papers according to the similarities among the paper's and researcher's features, in addition to finding neighborhoods of researchers with similar interests and criteria based on their ratings of the papers. GroupLens developed TechLens+ to test different hybrid algorithms [22]. The objective was to evaluate the performance of each used hybrid algorithm. The authors found that some hybrid algorithms perform better in recommending different kinds of papers, and the users' level of expertise affects their perception of the recommendations. Papyres [5] first uses CBF to find relevant papers. Then it refines the list using CF. Papyres uses explicit user ratings for 10 different paper qualities (i.e., originality, readability). This approach helps find papers with certain qualities (i.e., more readable papers or more well-organized papers) and at the same time find other papers that are highly rated by other researchers who have similar ratings in the past with the active researcher. Another approach applied CF based on user's tags of papers, then filtered the results based on text similarities of the papers [108]. [109] combined CBF and CF by applying topic modeling using text analysis and matrix factorization. Matrix factorization is a latent factor model, which is more successful than the neighborhood method.

PubRec [67] is an RS that suggests the most related papers for a particular paper in the researcher's library from the list of papers present in other researchers' libraries. The similarity of papers is measured according to similarities among the researcher's tagging of papers between the active researcher and other researchers s(he) is connected to. PubRec exploits social relationships and the trust of researchers. The trust in a researcher is not an absolute measure, but it is based on the knowledge of the researcher in the common topics of interest that s(he) shares with the active user (who needs recommendation). For example, a researcher will be more trusted if s(he) has more papers that belong to the topic for which recommendations are needed.

Other hybrid approaches, such as Scienstein [21] and Synthese [110], use citation-based methods to find similar papers. Scienstein combines CBF with CF and uses three ways of implicit rating: the citations data, authors' data, and monitoring users' actions (i.e., annotation, highlighting text, downloading or printing the paper, sending it to a friend) to infer the user's preferences. The system also allows users to rate the papers explicitly and to tag using different tag categories. Synthese is another hybrid RS that uses a citation matrix, but it uses PageRank values instead of a binary value.

Some other hybrid approaches monitor the behavior of users toward the papers. For example, some RSs infer the similarities among papers and among users using the users' viewing history, for example, QuickStep RS [4]. Users' interest neighborhoods are built by monitoring their digital libraries, assuming that when users download papers to their libraries, they implicitly rate these papers. It is assumed that the user is interested also in all the topics that these papers cover, as proposed in the PubRec system [67]. The users who share the same papers in their libraries are assumed to share similar interests.

Using papers that are bookmarked or downloaded in the user's library is beneficial in the recommendation process since these papers show the interested topics to the user. Next section discusses using social bookmarking tools as a starting point of building a research paper recommender system.

2.4.4 Social Bookmarking Tools

Managing references is not an easy job for researchers; it takes time and effort. Therefore, many social bookmarking websites (e.g. CiteUlike[7]) are used to save the interesting resources (e.g., books, journal articles, conference papers, websites) that the researchers use for their research and share with other researchers. In addition, there are many social reference management tools, such as Mendeley[8], Zotero[9], and Papers[10] available for researchers to use that share the following functions: help researchers to discover new papers through showing the most recently published papers; import their reference metadata automatically; provide access to references/bookmarks from anywhere; organize, tag, and search references/bookmarks using tags; share references/bookmarks with others; enable researchers to choose the style of formatting for their references and produce the reference list; enable annotation of texts and encourage communication between researchers.

2.4.4.1 Importance of Social Bookmarking Data

Social reference management systems (which are social bookmarking tools) aggregate academic resources' metadata that include data about the resources (e.g., author(s), title, abstract, keywords) or collaborative data (e.g., tags, ratings), which are valuable to be fed to RSs to produce personalized recommendations. Some bookmarking tools integrate easily with others—for example, CiteULike and Delicious—to synchronize all users' CiteULike bookmarks to delicious.com. CiteULike enables users to export their library to other reference management systems such as Mendeley or EndNote[11] using standard file formats (e.g., BibTeX file). Furthermore, usage data of papers can also be aggregated to produce an internal, system-wide analogue of an impact factor of papers. For example, Mendeley shows usage statistics for each paper that indicate how important the paper is to the community. Statistics include how many users have added this paper to their library, to which research discipline the readers belong by interest, and the academic status of the users.

Researchers have started realizing that social bookmarking tools provide valuable data that were created by users of these systems in their interactions (e.g., tags and annotations) with the

[7] http://citeulike.org
[8] http://www.mendeley.com
[9] http://www.zotero.org
[10] http://www.mekentosj.com/papers
[11] http://endnote.com

systems. After the announcement of Mendeley's API Binary Battle[12] many applications have been built using Mendeley data. For example, ReaderMeter[13] is a web-based application that takes advantages of the papers' usage data in Mendeley to analyze the impact of publications by a particular author based on the number of readers for each of the author's publications. Another application [111] uses Mendeley readership data to find the number of readers for each reference in a paper, so the reader of the citing paper can decide whether to read that reference or not.

2.4.4.2 Recommender Systems in Social Bookmarking Tools

Since social bookmarking tools collect information about researchers' interests, usage data, and citation data, RSs can be naturally applied to personalize the recommendations. Some bookmarking websites start applying RSs to recommend some papers that could be interesting for researchers but are not in their libraries. For example, CiteULike asks users to add at least 20 items to their library in order to be able to give recommendations. It involves user-based CF that compares the user's library with other users' libraries to select articles from those libraries to be recommended to the active user. CiteULike also applies item-based CF, the item–item co-occurrence (IICO) method that compares items that appear together to select the most frequently occurring ones to add to the recommendation list. The user can accept each recommended item by adding it to her library or reject it. The feedback of the users regarding the recommended list improves the later recommendation.

Mendeley Suggest is RS for the Mendeley social reference manager. It has two recommendation types: related research and personalized recommendation based on the user's library. Related research uses the paper's information to find related papers using different techniques: topic modeling, BM25, tags, CF, and in-text citation metrics to find similar papers based on co-citing them in a third paper using the distance between their appearance in the same sentence or same paragraph. However, personalized recommendation is available only for premium users. RS uses different recommendation algorithms: item-based CF, user-based CF, text analysis, and others. The performance of Mendeley Suggest also can be improved by the users' acceptance or rejection of the recommended items. Zotero is also developing its own RS. It seems that the door is open for more development of social bookmarking RSs for academic use.

[12] http://dev.mendeley.com/api-binary-battle/
[13] www.dcc.ac.uk/resources/extemal/readermeter

Even though CiteULike and Mendeley deploy recommendation services in social bookmarking tools, they do not exploit the wealth of the social data in the recommendation.

2.4.5 Summary of Recommending Research Papers

Developing RSs for researchers is a challenging job since many factors have to be considered to produce a personalized recommendation list that helps researchers achieve their goals, increase their knowledge, satisfy their needs, and at the same time give them the ability to customize recommendations and give feedback on the results.

Findings from surveying 33 RSs are summarized in Table 2-4: 7 RSs use the CBF, 12 use CF, and 12 use hybrid approaches, and 2 use other approaches. More than half of these RSs were developed after 2009, which shows that this research field is gaining popularity. However, only 1% of these RSs are running systems (RS of Mendeley, RS of CiteULike, and bX). The remaining RSs were evaluated either using small-scale user studies or were just algorithms, tested using large datasets without implementing the algorithms in complete RS. In both cases, the testing was done only for measuring the accuracy of the predictions of the proposed algorithms; only 3 RSs take into account the user satisfaction. Furthermore, only 5 RSs out of 33 use social relationships between people to recommend papers based on what is liked by other researchers whom the active user is connected with. In all of these five studies, the social relations are initiated by users (explicit) and it has been shown that recommendation algorithms based on explicit social relations have low coverage. In addition, all of these studies focused on the performance measures. We believe that social relationships in the recommendations can be harnessed to deliver a better balance between performance and non-performance metrics, by finding alternative methods of connecting users socially to each other in a way to increase not only the accuracy, but also the user coverage and diversity. Last, even though recent studies have demonstrated significant results in terms of recommendation accuracy, only 3 RSs use multidimensional ratings (i.e. multi-criteria rating) to deal with different aspects of papers (e.g., reading goals, pedagogical features, readability). Extending algorithms to work with multidimensional rating may help to increase user understanding, control and satisfaction of recommendations.

Table 2-4: Comparison between 33 research paper RSs

#	RS	Recommendation approach	Evaluate what!	MC RS	Trust	Citation or usage-based	Social
1	QuickStep [4]	Hybrid	Effectiveness of the algorithm	-	-	Usage-based	-
2	[112]	Hybrid	Effectiveness of the algorithm (precision and recall)	-	-	-	-
3	TechLens [2]	CF	- Measuring user opinion of the effectiveness of the algorithms. - Overall user satisfaction, and user opinion of the usefulness of the resulting system	-	-	Citation-based	-
4	[90]	CBF	Effectiveness of the proposed algorithms	-	-	-	-
5	TechLens+ [22]	Hybrid	- Accuracy and quality of recommendation - Users' perceptions about the quality of the recommendations	-	-	Citation-based	-
6	SERF [3]	CF	Users' search efficiency and effectiveness	-	-	Usage-based	-
7	Recommender of Papits system [95]	CBF	Effectiveness of RS	-	-	Usage-based	-
8	[97]	CF	Influence of trust-weighted reviews on document recommendations	-	Yes	Citation-based	-
9	[104]	CF	Comparing co-citation and co-downloaded methods in terms of quality	-	-	Usage-based	-
10	[106]	CF	Compare three different CF methods for usefulness	-	-	-	-
11	Zhang et al. (2008) [108]	Hybrid	Evaluating the influence caused by the size of neighbor user set	-	-	-	-
12	Papyres [5]	Hybrid	Test and compare five different approaches of defining user's neighborhood in terms of accuracy	Yes	-	-	-
13	ScienStein [21]	Hybrid	The services in the system is introduced, but there is no evaluation described	-	-	Citation-based and usage-based	-
14	Synthese [110]	Hybrid	Effectiveness of RS	Yes	-	Citation-based	-
15	[113]	CF	- Examine the effect of considering pedagogical elements in making recommendations - User satisfaction	Yes	-	-	-

#	RS	Recommendation approach	Evaluate what!	MC RS	Trust	Citation or usage-based	Social
16	bX [114]	CF	Running system	-	-	Usage-based	-
17	CiteUlike	CF	Running system	-	-	-	-
18	[92]	CBF	Effectiveness of proposed algorithm	-	-	Citation-based	-
19	He et al,, 2010 [89]	CBF	Measure the recommendation performance	-	-	Citation-based	-
20	[107]	CF	Effectiveness of proposed algorithm	-	-	-	-
21	[70]	Hybrid	Effectiveness of hybrid approach of CF and group collection, group membership in comparison to each one separately	-	-	-	Yes
22	PubRec [67]	Hybrid	Effectiveness of proposed algorithm	-	Yes	Usage-based	Yes
23	Wang and Beli [109]	Hybrid	Comparing the performance of three CF techniques	-	-	-	-
24	[94]	CBF	Effectiveness of proposed algorithm	-	-	Citation and co-authorship graph	-
25	[12]	Hybrid	Effectiveness of hybrid approach of CF and watching network in comparison to each one separately	-	-	-	Yes
26	[88]	CBF	Effectiveness of proposed algorithm	-	-	Citation-based	-
27	Mendeley	CF	Running system	-	-	-	-
28	DOCEAR [115]	CBF	Compare the performance of different variations of CBF	-	-	-	-
29	[116]	Different algorithms	Compare the performance of different algorithm and fusing of them to test the social recommendation algorithm	-	-	-	Yes
30	PReSA [69]	Hybrid	Efficiency and effectiveness of proposed algorithm	-	Yes	Usage-based	Yes
31	PaperTaste [99]	CF	Effectiveness of proposed algorithm	-	-	Citation-based	-
32	[100]	Greedy Clique Expansion Algorithm + PaperRank Algorithm	Effectiveness of proposed algorithm	-	-	Citation-based	-
33	[101]	CF	Performance of the proposed algorithm	-	-	Citation-based	-

CHAPTER 3: MULTI-CRITERIA RATINGS FOR RESEARCH PAPER RECOMMENDER SYSTEMS

As discussed in the previous chapter, to alleviate the bias in the single rating, multi-criteria rating recommenders were proposed. Using different rating criteria to evaluate the quality of the items provides more data about the rated items and also about the users' preferences, which can enhance the collaborative filtering performance. Finding users who are similar to the target user considering a certain aspect of the quality of the item allows to recommend items (e.g. research papers) from similar users who have similar ratings to the target user with respect to that aspect and this can improve the recommendations, and possibly, the user satisfaction.

This chapter is dedicated to answer the following question:

> ***RQ1:*** *How do users perceive multi-criteria rating recommendations?*

This broad question is divided into the following specific research questions:

> **RQ1.1:** What are the most important rating criteria in evaluating a research paper?
>
> **RQ1.2:** What are users' preferences about using overall ratings versus multi-criteria ratings?
>
> **RQ1.3:** Do users prefer to have control over the importance weights of multi-criteria ratings during the recommendation process?
>
> **RQ1.4:** Are the criteria domain-dependent?

We conducted a qualitative study to find answers to research questions: RQ1.1 to RQ1.3. Then, we confirm the rating criteria ranking using quantitative study as well as confirming our finding for RQ1.3. We also generalized the findings with two research areas using the quantitative study.

3.1 Qualitative Study

Most CF algorithms require users to give just one overall (global) rating and then use the averages of all users' ratings to correlate the items (or users) and compute "neighborhoods". This approach is straightforward but not flexible enough to provide adequate details about the quality of the rated item/service. The inflexibility of global ratings produces biased recommendations because two users may give the same global rating from two different perspectives [5]. For example, two researchers may rate a paper the same, but the first researcher's evaluation is based on the paper's readability while the other's is on the paper's novelty. For this reason, some RSs are based on CF algorithms or hybrid approaches that use multi-criteria ratings based on two or more perspectives (dimensions). However, in all the existing MCRS, the rating criteria are chosen by researchers and do not allow the users to change the importance weights of these criteria.

In this study, we investigate researchers' opinions of the most important criteria in rating the quality of a paper (RQ1.1). The important decisions include how many and which criteria to include, whether users prefer to consider the overall rating or multi-criteria ratings (RQ1.2) whether to include the possibility of users' assigning different weights to the criteria (RQ1.3) and how the user interface for ratings should be organized. There is a danger in giving more control to the user and increasing the complexity because that may lead to cognitive overload and reduce the user's rating activity. Although all published research papers are peer-reviewed and evaluated for their significance and novelty, we are looking to the users' reviews as consumers for these papers. The importance of that is to try to develop a recommender system that take into account the user's preferences using the multiple ratings and the user's weights for each rating criterion. We argue that different weights can produce a more personalized recommended paper list to each user.

Recent studies of user-control in the context of recommender systems showed that the relation between the user-control and satisfaction is affected by the knowledge level of the user [117] and the user's interests [118]. It also shown that there is a relation between the user control and user engagement in the recommender system and the user characteristics affect the user engagement and the user experience [119] [120]. [121] showed in the TasteWeight recommender system for music that using the visualization allowed bigger interaction and explainability and enhanced the general user experience. In our work, we want to investigate the users' opinion in giving the user control on the importance weight of different rating criteria. Giving different weight for each criterion is expected to help in the personalization of the recommendation and also in increase the user satisfaction of the recommender system. The output of this study is expected to help the designer of the recommender system interface to understand the user's needs.

Following the framework proposed in [122]) we concentrate on two layers of the model: deciding upon adaptation (DA) and applying adaptation decisions (AA) to elicit the users' opinions. We use a qualitative method by applying focus group discussion to involve the end users and their perspective throughout this study.

3.1.1 Methodology

In RS design and development, quantitative research methods are commonly used to evaluate the accuracy, efficiency and the effectiveness of proposed algorithms as well as user satisfaction [5][95]. However, the qualitative method is well suited to exploring different options and user requirements. To our best knowledge, the question of which features (i.e., criteria) to consider in evaluating the quality of a paper has not been addressed in a systematic way, and all existing approaches that include multiple rating criteria seem to have been developed based only on the authors' intuition. We adopted a focus group approach, a qualitative research method suited for exploratory research that would allow us to examine 1) whether the participants thought including different ratings criteria is useful in finding higher quality research papers and 2) whether they support the idea of having the user control the weight of different criteria. We also want to extract and confirm some guidelines for user interface design that can make the RS more understandable and user-friendly.

A focus group, as defined by [123], is a moderated discussion on a predefined set of topics with 6–12 participants. The advantage of having small group discussion is allowing participants to develop shared understanding of the topics while voicing their opinions. The process uses open-ended questions to enrich the discussion with different ideas, perspectives and conflicts that reveal the similarities and differences in participants' thoughts. Some questions should be defined before conducting the focus group, whereas some may evolve as follow-up questions based on the participants' responses. One of the big advantages of a focus group as an exploratory method is its open-endedness, the fact that the direct interactions with and among the participants enable exploration and redefining the scope of ideas.

3.1.2 Participants

Eight people participated in this study: one postdoctoral fellow, two Ph.D. candidates in their fourth year of research, and five master's students in their second year of studies. The participants were recruited from the Computer Science Department of the University of Saskatchewan. We invited these people specifically because we aimed to engage active young researchers who still need to read research papers for their studies and for writing their theses. Graduate students usually need to read dozens of papers to be familiar with their research area, to find problems that they can contribute to solving and, finally, to write their theses. Graduate students usually struggle to find good papers to read when they are starting their studies. Participants were invited personally through e-mail. Eleven of the invited people agreed to participate in the study, but three could not participate because of other commitments. The participants' demographic data are shown in Table 3-1. Most of the participants were international students who had studied previously in different countries, which meant their diverse educational backgrounds could enrich the discussion.

Table 3-1: Participants' Demographics Data

Participant #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Degree	Master	Postdoc	Master	Ph.D	Ph.D	Master	Master	Master
Gender	Male	Female	Female	Female	Male	Male	Male	Male
Country of Origin	Iran	Iran	Canada	Nigeria	Nigeria	Taiwan	China	Canada

3.1.3 Focus Group Settings

The focus group was held in a quiet room in the Social Sciences Research Laboratories (SSRL), University of Saskatchewan, dedicated to conduct qualitative studies (e.g., focus groups or think-aloud sessions). Preparation of the room included providing microphones for audio recording the session and a round table with comfortable chairs to enable the participants to communicate with each other easily.

When the participants arrived, they were welcomed and asked to read carefully and sign the consent forms (please see Appendix B.1, Ethical certificate attached in Appendix A). They were reminded that the focus group discussion was audio-recorded to make it easy to transcribe the session without missing any of the participants' opinions. They were also told to avoid interrupting any of the other participants so that all participants had the chance to complete their ideas. This also helps the transcription by preventing the audio file from being garbled.

3.1.4 Questions Asked

Two types of questions were asked during the focus group session. First, we asked closed questions to examine how familiar the participants were with RSs:

C1. Have you been recommended items, or are you familiar with any kinds of recommendations?

C2. Are you familiar with research paper RSs?

Then we asked open-ended questions as follows:

O1. When you want to read a paper, how do you choose that paper? What criteria do you consider?

O2. In using a RS for research papers, do you prefer to see overall ratings (one rating as general rating) for the paper or multiple criteria? Justify your answer.

O3. If you are asked to evaluate and rate a research paper, in your opinion, what are the three most important rating criteria?

O4. What do you think about giving the user the control to assign importance weights to those criteria?

3.1.5 Analysis and Results

The discussion began with questions C1 and C2. We noticed that the participants were very motivated to talk about the topic. All of them were familiar with RSs, specifically in recommending items to buy; seven of them were familiar with RSs used by scholarly websites, such as Academia.edu and Google Scholar Updates. Below are some direct quotes from participants as they expressed their hypotheses about the recommendation algorithms used:

- Participant 2 (P2): "I think they [Google Scholar Updates] may use some index [keywords] that you use in your papers. Try to match it and it will suggest to you some papers that have been recently published in your field."
- P8: "Google Scholar[14] seems to list them in order of how many citations they have. That's the only thing I've noticed on Google Scholar."
- P1: "I think Google Scholar itself is somehow [a] recommender system based on [the] number of citation[s] that the paper has? For example, if you search using keywords and you see a paper that has thousands of citations, so you prefer that one to the others." P5: "Yeah,

[14] http://scholar.google.ca/

but they [Academia.edu] just send me an e-mail with publications. So maybe they use the area of interest. I don't know."

- P4: "When you sign in to academica.edu it comments with recommendations so are you interested in this area? You either click [to] select it or not; once you selected it, it starts recommend papers in these areas. I was wondering one time that why it recommends some papers that I feel they are not in my research areas, I looked into my profile and I discovered that I've actually selected that before."
- P4: "ResearchGate[15] send e-mail by questions and their answers using the keywords that I used previously to search."

3.1.5.1 Brainstorming the Evaluation Criteria of Papers

Participants were asked to think about how they usually choose a paper to read and the reasons (i.e. criteria) that they choose to read a paper or not. They were asked to define each criterion so we would know exactly what they meant by each of them. They discussed the following 9 criteria.

1- **Clarity:** Participants converged toward defining clarity of the paper as how well the paper is written in English, whether it is using simple terms and how the style of writing makes the ideas presented by the paper obvious. They thought this is one of the criteria that make them decide to read the paper or not because the clearer the paper the more accessible and understandable it is. However, participants distinguished between two terms: clarity and understandability. They referred to clarity mostly as the quality and simplicity of the language used, while understandability was to mean how clear and logical the paper is in presenting the technical approach and the research methods used. For example, P5 said, "I think one is about grammar... and it was written poorly. So that might make it difficult to read. The second one is the paper is written well in English but it's still difficult to get what that person is talking about, but I think that might be technical clarity."

2- **Technical clarity:** participants felt that the technical content of the paper is one of the most important criteria in technical disciplines, such as engineering and computer science. However, the technical content may not be sound or may not be presented in an

[15] http://www.researchgate.net/

understandable way. For example, some authors use a lot of difficult technical terms that make the paper hard to understand. Similarly, complex mathematical formulas may be introduced without giving a hint or justification of why they are necessary. In some empirical or applied science papers the description makes it difficult to understand what the goals of the experiment are and why a specific evaluation or statistical evaluation method is chosen, assuming that the reader is very familiar with the subject.

3- **Willingness to Cite This Paper:** this criterion corresponds most closely to the general opinion of readers regarding liking or disliking the paper, if they intend to cite this paper in the future or not. This means they like it and find the information useful for their work. A high rating on this criterion would mean that, even though users may find the paper not useful in some aspects or may not like its clarity, it still brings value, can be useful and is worth citing.

4- **Length of the paper:** some participants thought that one of the criteria that encourage them to read the paper is its length. The shorter the paper, the higher the chance it will be read. "Sometimes a paper's, like, 40 pages and you're, like, I probably don't want to read this," P3 said. However, they thought this could be a ranking criterion rather than a rating criterion, since the page length is an objective feature rather than a feature requiring a subjective evaluation (rating).

5- **Closeness to purpose or task:** participants defined this criterion as how much the paper is relevant to what the user is doing at the moment (e.g. writing a paper, finding another approach to solve the same problem). For example, if users read a paper, they might want to read another similar paper that uses a different approach or different method or frames the same problem from a different perspective. P2 commented that "Sometimes [the user] want[s] something newer. Sometimes [the user] want[s] something close." This criterion was raised by one of the participants and did not receive support from the others.

6- **Relevance:** the participants defined this criterion as the similarity between the content of the paper and the user's research interests. The recommended papers must be relevant to the user's interest somehow. The participants thought that this criterion is more important if the user has interests in multidisciplinary research fields. They wanted to be able to judge the relevance even though this can be automated (e.g. using CBF). They also suggested enabling the user to tag papers so that they can relate the paper to topics according to their own choice.

7- **Comprehensiveness:** participants described a paper as comprehensive if it reviews what has been done in the research area (i.e., literature review) without going deeply into detail. Some participants called this criterion "coverage." They mentioned that comprehensive papers are more important for new researchers that need to read "something that at least gives a brief overview… a summary of what has been done like literature review", P4 said.

8- **Difficulty level:** some participants insisted that the difficulty level of the paper is important to consider in deciding to read the paper, but they could not agree on what defines a paper as difficult to read. While there are linguistic metrics of text complexity that could be used, the participants felt that the problem with the difficulty of research papers is more nuanced and text complexity metrics are not sufficient to evaluate it. There was a strong conflict in the discussion of this criterion. Some participants argued that the paper could be difficult because it has many mathematical equations, theorems and proofs. Other participants said the mathematical issues were only in the applied sciences, but when we consider other fields such as social sciences, the papers do not usually contain any mathematics yet can still be difficult to understand. They recognize that some readers are good at mathematics and may feel that papers with lots of mathematics are easy. Papers that use many technical phrases, acronyms or jargon that are specifically used in that research field are difficult to understand to those who are unfamiliar with them. One participant commented, "Sometimes the paper is very hard to read. I don't mean English but in the technical field. They use lots of English phrases that are specifically used in that field." Depending on the reader's preparation or experience, the paper may be easier or harder to read. Some other participants found that papers that discuss experiments or user studies are easier to read than the papers that describe theories. "But I think that is also subjective. It depends on the kind of research that the paper discusses. Like, if it is strictly theoretical research then it might be difficult for most people to research. But if it is, like, experimental… it is easier to understand," said P4.

Most participants stated that the distinction between theoretical and experimental papers is important but not available in any of the existing research paper bookmarking websites, such as CiteULike[16], Mendeley[17] or Zotero[18]. P2 said, “If I want to trust or rely on any RSs, I want to see it ignore the paper that comes with lots of theoretical stuff.” This distinction can be done by enabling the user to categorize papers as experimental or theoretical. Then users can specify in their profile that they prefer one of them to the other (e.g., experimental), which can be another feature to be used by the content-based filtering or hybrid RS to find matches.

9- **Value added by reading the paper.** This criterion refers to how much new information the user gains by reading this paper. Some participants thought that it is useful to rate the usefulness of the content of the paper based on their knowledge, whereas other participants thought that it would be hard to rate old papers. P8 said, “I feel like if a paper’s really original and novel then it would be easier to give it a higher value added rating because you can see how many things you can do to build on this paper. But older papers you might say it seems like this old idea that you can’t build off of.” P5 commented that an older paper could not add value for them because they have already read many and newer papers in the same area.

3.1.5.2 Other Criteria Mentioned in the Discussion

The participants mentioned some other criteria that are useful in evaluating a paper, such as how recent the paper is, the authors’ reputation (h-index) and the number of citations that the paper has received. They discussed the importance of the year of publication and the relationship between the publication date and the number of citations, acknowledging that it is not straightforward to base a conclusion about how important the paper is on its citations. P4 said: “The paper might be new. It has not had much citation, but it is very good.” The reputation of the author is also important; users may be interested in reading a paper written by author(s) that they know and follow. However, the participants realized that, although these criteria are important to consider, it is better to consider them as ranking criteria for the results rather than rating criteria for users, since all of these criteria have numeric values that are usually fixed at the time of

[16] http://www.citeulike.org/
[17] http://www.mendeley.com/
[18] https://www.zotero.org/

considering them (e.g., the length) or change slowly (e.g., the number of citations, the year of publication, the h-index of the author).

After we combine the explicit and implicit users' interests and the paper ratings and the content of the paper and using a hybrid approach, the objective criteria (e.g., paper length, text complexity, year of publication, authors' reputation, and number of citations) can be used to narrow and order the list of results.

3.1.5.3 Overall Rating versus Multiple Ratings

When we asked the participants Question O2 (whether they preferred to have the overall rating of the paper or multiple ratings based on different specific criteria), the participants expressed wishes for an overall rating and ratings based on three specific rating criteria for two reasons. First, the user sometimes strongly likes some aspects of the paper, but their general opinion of it may be low. All the participants agreed with the opinion of P5: "You still give your opinion on the different dimensions. But generally you still come up with the overall impression whether you accept or reject [the paper]." Second, the user might have different opinion from the available criteria for the ratings and that can change the overall rating of the paper.

The participants thought it is important to have different criteria that reflect different aspects of the paper's quality, which would make the recommendations more personalized to their needs. All participants agreed that the obvious disadvantage of having multiple ratings for each paper is that "It's too much work to click on all the rating criteria for each paper." However, the solution is to "keep [the ratings] short and simple," design the system carefully and display the benefits of having multiple ratings on recommendations that they receive.

3.1.5.4 Choosing the Most Important Three Rating Criteria

After the participants felt that there were no more criteria to add to the list, we asked them to choose the three most important criteria among those produced in the brainstorming discussion related to O1. We initiated the discussion with questions O3. One participant suggested starting by removing the least important criteria and focusing later on ranking the remaining ones. First, most (five or more) of participants suggested removing the "willingness to cite the paper" and "value added by reading the paper" because these criteria expressed a general impression about the paper that could be replaced by the overall rating. The participants thought that the "length of the paper" should be removed because it is an objective, static parameter that could be used by the RS as a ranking criterion after the set of possible recommendations was generated. Another criterion that participants felt that could be automatically inferred by the RS, instead of relying on user rating, is the "closeness to research area." Each user has a user model that contains the user's preferences, the research interests and any related topics of interest. The user model is built automatically using data that are entered during user registration or collected by the system based on the user's history of searching, bookmarking, etc.

Table 3-2 shows the resulting ranking of the rating criteria. All participants agreed that the clarity of the paper is the most important rating criterion. It is important for the users to read well-written papers that use simple and well-defined terms and that are easy to read. One participant commented, "number one should remain clarity and that it is easy to read," expressing the group consensus. However, there were some comments that suggested more nuanced views. Another participant pointed out that a paper's clarity is related to its difficulty and that the difficulty criterion is subjective. Yet difficulty came only fourth in the ranking and was supported by only three of the participants.

Technical clarity became the second most important criterion, as ranked by seven participants. The paper is technically clear when the technical details are presented so that the work can be reproduced. Such details include consistent mathematical notation and formularization, detailed description of the research method and the experimental setup.

Table 3-2: Ranked List of the First Five Important Rating Criteria

Criterion	**Ranking position**	**Number of users agreed on the ranking position**
Clarity	1	8 (100%)
Technical clarity	2	7 (87.5%)
Relevance	3	5 (71.4%)
Difficulty level	4	3 (37.5%)
Comprehensiveness	5	2 (25%)

There was a long discussion to choose the third-ranked criterion among relevance, difficulty level and comprehensiveness. Finally, in a vote taken to resolve the question, five participants chose relevance.

3.1.5.5 Control Over Different Rating Criteria

When the participants have been asked O4, they showed their strong support of the idea of enabling the users to show their preferences for different rating criteria by assigning weight to the criteria. They felt that the ability to change their preferences could help users find better papers according to the criteria that they prefer most or to a combination of different criteria. They acknowledged that users' needs change with time, such as when senior researchers might give more weight to the technical clarity, while new researchers might give more weight to clarity and comprehensiveness. The participants discussed some ideas about how the user interface could enable user control over the weights of the criteria which are discussed in the next section.

3.1.5.6 Paper Recommender System and User Interface Design Implications

The focus group discussion proposed some interesting ideas for designing the paper RS shown in see Table 3-3. Participants showed a preference for having sliders to adjust the importance weight for each of the criteria with a default setting for each one (e.g., in the middle of the slider). However, they mentioned that it is important to make a minimum threshold value to not allow the user to ignore completely any of the criteria. For example, users would not be able to make the slider go less than 20% for the clarity criterion.

Table 3-3: Implication of Raised Issues on the Paper RS design

Issue or Aspect	Effect on the RS Design
Topic relevancy to the user's interests	Use hybrid recommendation approach based on model of user interests and ratings.
User control over the weights of the rating criteria	Use sliders for each criteria; use checkboxes to choose which criterion the user wants to change the weight; or enable ranking the criteria
Overall rating or multiple ratings	Use both overall rating and multiple ratings; overall rating is used to reflect the general impression of the user about the paper
Experimental or theoretical papers	Enable users to classify the papers according to predefined categories (e.g., ACM classifications) and specify preferred categories in their profile
Users do not like to fill in much information	Keep criteria few, well defined and show tips that give clear descriptions
Users cannot do the multiple ratings before reading the paper	Enable multiple ratings only for users who have read the paper (e.g., by using checkboxes)
Author reputation, publication date, number of citations, paper length, text complexity (measured with text analysis tools as one component of the difficulty level of the paper)	Recommendation algorithm to use such "static" parameters as ranking criteria for ordering the recommendations
Relate papers to topics and evaluate the paper relevancy	Enable tagging

One of the participants suggested ranking the three criteria instead of giving weight value by using sliders, "and just put this is the first one, this is the second one, and the third one… that would be way easier." Another participant suggested enabling the user to choose three criteria out of five, for example, by selecting them in a checkbox, and then the user could change the weight for the selected criterion (or criteria). He commented, “Maybe it should be dynamic.”

3.2 The Quantitative Study

In the previous section, we discussed the focus group experiment that aimed to extract the most important criteria that user consider when reading a research paper. However, the study was conducted with small group of users because the focus group could not be done with a large group. In addition, the focus group was conducted with users from the same discipline: Computer Science. In order to generalize the results that we got, a quantitative study was designed. There were two main objectives for this study. First, we wanted to validate the three most important criteria to be included in a multi-criteria RS that the users chose in the focus group. Second, we wanted to check if users from other discipline agree with the same ranking of criteria as the Computer Science users or not.

3.2.1 Methodology

An online questionnaire was used in this study. Participants, who were graduate students or postdoctoral fellows, were recruited for this study by contacting them personally, or sending emails. The participation in the study was completely voluntary. The details of the participation in the study along with link to the online questionnaire were sent by the email. Any participant had to agree to the consent form, attached in Appendix B.2, that describes the process of the participation before they proceed to the questions. There was no need to fill any personal information that lead to their identity. Each form was given a number to ensure that participants are anonymous.

3.2.2 Participants

The questionnaire was available for ten days, 102 participants started the questionnaire, but 71 participants completed it. Two of them did not consent to participate in the study. Therefore, 69 participants were included in our study. Out of 69 participants: 28 were females and 41 were males. Their ages were: 13 participants in the age group of 21 to 25, 29 in the 26-30 age group, 22 in the 31-35 age group, 4 in the 36-40 age group and one participant was above 40 years old. Figure 3-1 and Figure 3-2 show the percentages of the participants according to their gender and age group respectively.

The analysis of the participants' data according to their research disciplines show that 40 participants were from the area of Computer Science and 28 participants from Pharmacy and Nutrition. Based on their graduate studies: 37 participants were master students (22 computer science, 15 pharmacy and nutrition), 28 PhD students (16 computer science, 12 pharmacy and nutrition) and 4 postdoctoral fellows (2 computer science, 2 pharmacy and nutrition). Figure 3-3 and Figure 3-4 show the breakdown of the participants' data according to their discipline and their degree of study respectively.

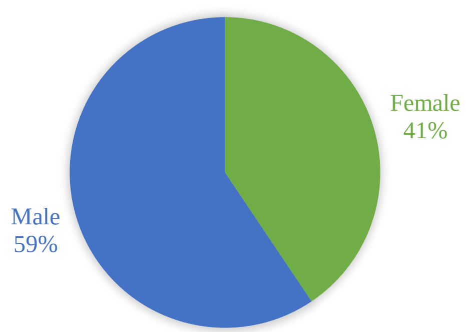


Figure 3-1: Percentage of participants according to their gender

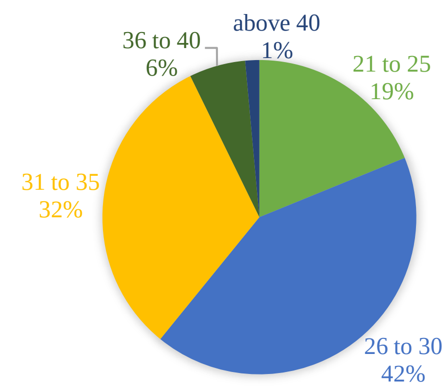


Figure 3-2: Percentage of participants according to their age group

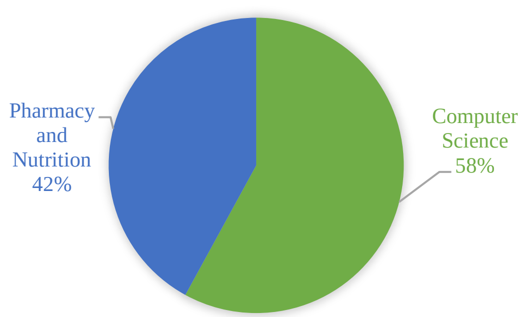


Figure 3-3: Percentage of participants according to their disciplines

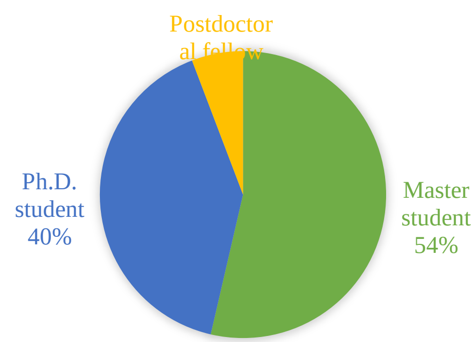


Figure 3-4: Percentage of participants according to their graduate studies

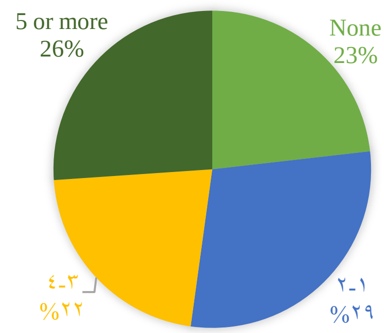


Figure 3-5: Percentage of participants according to their number of publications

3.2.3 Questions Asked

The questionnaire was short because the aim was to validate the questions/findings that have been asked/come up with in the focus group. First, we asked some questions to collect data about the users' demographic data. Then, we asked the participants about their number of publications so we could infer their experience in reading/writing papers, Figure 3-5 shows the percentages of participants according to their number of publications. Then, the main question is asked; participants asked to rank six rating criteria according to their relative importance from the users' point of view. These six criteria were the first most six criteria that the participants agreed upon in the focus group. The last question was an open-ended question: "*Do you prefer to set an importance weight to each rating criteria to get paper recommendation according to your settings? Please justify your answer*". The aim was to collect the users' opinions regarding the idea of giving some control to the user in the recommendation process. Open-ended question was suitable since we wanted to explore the users' ideas about the possible advantages and disadvantages of applying weighted criteria in our proposed recommender system.

3.2.4 Analysis and Results

The demographic data of the participants are discussed in section 3.2.2. Therefore, in this section, the ranking question and the open-ended question are discussed.

To analyze the users' answers of the ranking question, we calculated the total score of each criterion, which is based on a weighted calculation described below. For each criterion, the number of users who chose it at a specific rank is counted, then for each rank, the counter is multiplied by the weight. For example, if we want to calculate the total score for one criterion, since we have 6 criteria, the maximum weight, which is 6 in our case, will be multiplied by the number of participants who chose this criterion as the first important one, the second weight which is 5 will be multiplied by the number of participants who chose this criterion as second important criterion in their ranked list, and so on. Then the multiplied values are summed to produce the total score for this criterion. This process is repeated for all six criteria, and then the total scores are sorted in descending order to produce the ranked list of the criteria according to all users. Table 3-4 shows how the total score is calculated.

Table 3-4: The way of calculating total score of specific criterion using weighted calculation

Total responses		**Weight**		
Rank 1 count	X	6	=	
Rank 2 count	X	5	=	
Rank 3 count	X	4	=	
Rank 4 count	X	3	=	
Rank 5 count	X	2	=	
Rank 6 count	X	1	=	
			+	Total score for specific criterion

3.2.4.1 Criteria Ranking Discussion

First, the ranking for the criteria were calculated for all users. Table 3-5 shows the ranks with the total score for each criterion. When comparing how the participants in this study chose the first four criteria, we can see that they chose the same first three criteria as the participants in the focus group study. However in the focus group, participants agreed that the criterion "value added by reading the paper" can be considered as an overall rating instead of one of multi-criteria ratings. So, we removed this criterion from the comparison in the focus group. Since the objective is to choose the first three important criteria that all participants agreed upon, we can conclude that the chosen criteria from both groups are: relevance, value added by reading the paper, technical clarity, and clarity.

Table 3-5: Ranking of the rating criteria for all participants

Rank	**Criterion**	**Total score**
1	Relevance	355
2	Value added by reading the paper	307
3	Technical clarity	269
4	Clarity	240
5	Willingness to cite the paper	180
6	Difficulty level	98

We also calculated the different ranks of the criteria for different groups of participants. For example, we calculated the ranks based on the participants' major of study, based on the degree that they were pursuing, and based on the number of publications as an indicator of the previous knowledge of the participants in reading/writing research papers. First, we distinguished the rankings between participants who were studying computer science and participants who are studying pharmacy and nutrition. Table 3-6 and table 3-7 show the criteria ranking of different groups. Participants in the two groups chose the same first four criteria with also the same order except for the last two. Computer science students chose technical clarity to be the third criterion and the clarity of the paper to be the fourth while pharmacy and nutrition students chose the opposite. However, the total scores for the clarity and technical clarity criteria in the case of pharmacy and nutrition students were almost the same.

Table 3-6: Ranking of the rating criteria for computer science students

Rank	Criterion	Total score
1	Relevance	202
2	Value added by reading the paper	183
3	Technical clarity	163
4	Clarity	132
5	Willingness to cite the paper	104
6	Difficulty level	56

Table 3-7: Ranking of the rating criteria for pharmacy and nutrition students

Rank	Criterion	Total score
1	Relevance	153
2	Value added by reading the paper	124
3	Clarity	108
4	Technical clarity	106
5	Willingness to cite the paper	76
6	Difficulty level	42

The criteria ranking was also compared based on the degree pursued. Tables 3-8, to Table 3-10 show the ranking of the criteria of PhD students, master students and postdoctoral fellow respectively. As noticed from the tables, PhD students and postdoctoral participants have identical ranks while master students are different from them. First, master students chose in the first place how much value the paper added to their knowledge. They were focusing on getting familiar with the research area by getting new ideas from reading the papers that add new concepts to their knowledge. In addition, they gave higher rank to the clarity of the paper in comparison to technical clarity. It seems that new graduate students prefer to read papers that are well written and easy to read rather than papers focusing on the technical details.

Table 3-8: Ranking of the rating criteria for PhD students

Rank	Criterion	Total score
1	Relevance	139
2	Value added by reading the paper	124
3	Technical clarity	120
4	Clarity	94
5	Willingness to cite the paper	90
6	Difficulty level	36

Table 3-9: Ranking of the rating criteria for master students

Rank	Criterion	Total score
1	Value added by reading the paper	162
2	Relevance	154
3	Clarity	135
4	Technical clarity	115
5	Willingness to cite the paper	94
6	Difficulty level	57

Table 3-10: Ranking of the rating criteria for postdoctoral fellow

Rank	Criterion	Total score
1	Relevance	22
2	Value added by reading the paper	21
3	Technical clarity	14
4	Clarity	11
5	Willingness to cite the paper	11
6	Difficulty level	6

Then the criteria ranking were analyzed based on the participant's level of knowledge of reading papers. We consider the number of publications as another evidence of the participants' level of knowledge of reading papers. There are 16 participants who do not have publications, 20 participants have 1-2 publications, 15 have 3-4 publications and 16 participants have 5 or more publications. Tables 3-11 through table 3-14 show the ranking of the criteria based on the number of publications that the participants have. As shown, all of the participants chose the relevance to be in the first place, followed by "value added by reading the paper" in the second place for participants who have 4 published papers or less. However, when the reader of the paper has more experience in writing papers, they focus more on the technical value of the paper, as shown in table 3-14. All participants gave higher rank for technical clarity except the participants who did not publish any paper, they focus more on how well the paper is written.

Table 3-11: Ranking of the rating criteria for participants who do not have publications

Rank	Criterion	Total score
1	Relevance	80
2	Value added by reading the paper	67
3	Clarity	66
4	Technical clarity	58
5	Willingness to cite the paper	39
6	Difficulty level	30

Table 3-12: Ranking of the rating criteria for participants who have 1-2 publications

Rank	Criterion	Total score
1	Relevance	112
2	Value added by reading the paper	92
3	Technical clarity	77
4	Clarity	60
5	Willingness to cite the paper	60
6	Difficulty level	24

Table 3-13: Ranking of the rating criteria for participants have 3-4 publications

Rank	Criterion	Total score
1	Relevance	76
2	Value added by reading the paper	67
3	Clarity	57
4	Technical clarity	55
5	Willingness to cite the paper	36
6	Difficulty level	24

Table 3-14: Ranking of the rating criteria for participants who have 5 or more publications

Rank	Criterion	Total score
1	Relevance	92
2	Technical clarity	79
3	Value added by reading the paper	76
4	Clarity	59
5	Willingness to cite the paper	47
6	Difficulty level	20

3.2.4.2 *Analysis of the Open-ended Question*

We asked one open-ended question at the end of the questionnaire to collect the participants' opinions about liking/disliking the idea of allowing users to control the importance weight of each criterion. 60 out of 69 participants (86.95%) stated that they would like to control the importance weight of the criteria by changing the weight of each one. First, the opinions of the participants who did not agree with the idea will be discussed, followed by the opinions of participants who liked the idea.

Participants who answered "No" for the open-ended question gave different reasons. Three participants did not justify their answers. Two pharmacy master students (P81, P82) mentioned that their criteria of choosing papers to read are changing by time due to changing in their research objectives. One computer science master student (P117) thinks that there is no need to control the importance of the rating criteria and the recommender system "should be smart enough to work with only implicit feedback". One pharmacy PhD student (P90) raised an interesting point that is related to using recommender systems rather than using different rating criteria. He wrote that even though the recommender system might accelerate finding papers relevant to the user's interest, some papers that are interesting but not very relevant might be excluded from the list of recommended papers. He considered searching for a paper using a search engine as a learning process so the user can find some very relevant papers and some interesting papers but not completely relevant, and he thinks that the user might miss this opportunity by using recommender system. Another participant (P115, a computer science PhD student) similarly stated that the user might miss some papers by using weights that do not fit the user's priorities but the user still wants to see them. P47 was a PhD computer science student who thought that it was difficult for users to evaluate the importance numerically.

Participants who liked to have the control on the importance weights of the rating criteria described it by very positive words such as "great", "smart idea", "logical", "would be amazing", "convenient", "helpful". However, eight participants did not justify their answers.

Most of the participants who agree with the idea thought that using a multi-criteria rating recommender system flexible with respect to the importance weight would save their time and efforts and make them become more productive.

The reasons that the participants mentioned about why they like to have some control were:

1. Control over different criteria would allow them to cope with the change in their objectives and needs over time. P94 commented: “sometimes I change my mind about the factor that is most important to me at a certain period of time.”
2. Recommender system could help new researchers who struggle in finding papers to read since they are new in the research field, and it would be useful and helpful if users were able to specify the criterion that they want to focus on at a certain time.
3. Some criteria were more important than others, and the system “should allow different settings”, P101 commented.
4. Users preferred to control the available various options. P56 said: “when facing multiple options, I prefer to have the option to weight my answer.”

Searching for papers would be easier since providing the system with the user’s options regarding the rating criteria would narrow the search results. P72 commented: “the control on the rating criteria will help to rely on the user’s preferences and narrow the search instead of digging into somewhat unrelated stuff”

P87 added that the proposed idea is “great which I think most of the current search engines is missing”

Participants raised some design issues in their feedback. For example, they suggested that enabling the user to reorder the criteria according to their importance each time they need to get different recommendation based on the new order. Some other participants suggested using percentages to give each criterion the weight importance. For example, if the first criterion is the more important than the other two, the user can give 50% for it and the other 50% can be distributed evenly to the other two criteria. Participants also raised an important issue regarding the minimum value of the importance weight. They suggested to not allowing users to give zero weight for some of the rating criteria. For example, P92 said: “I do not want to waste time reading a paper that is not relevant to what I am looking for no matter how well written.” Therefore, the system should consider the paper relevancy in the recommendation process and not ignore it completely even though the user does not make the relevance criterion as the most important one. In contrast, another participant suggested not recommending a paper that meets the low preference of the user in one criterion despite the other rating criteria. P101 suggested that if a paper does not meet one of the criteria, it should not be recommended regardless of the results of other criteria.

Even though the feedback of the participant in answering the open-ended question was highly positive, there were some concerns that made them wondering about using the recommender system in general. For instance, some of them thought that by applying system filter they might miss some papers that were useful but not very relevant to their research, which could be captured by using search engine.

3.3 Reaching an Impasse

We planned to develop a recommender system that uses the three most important rating criteria discovered in the studies above, while taking into account the user's social relationships. However, we stumbled upon an impasse regarding the future evaluation of the approach. We were not able to find a dataset with user ratings using different rating criteria. Even if such a set became available in the future, most likely it would not have had exactly the same rating criteria that we discovered and wanted to test. The other option for evaluation was to generate multiple ratings for research papers synthetically. However, along the need to generate user ratings for the papers, and we there was also the need to generate social relationships between users. Using synthetic data to such an extent, would have made our results questionable, since according to [14]: "*drawing comparative conclusions from synthetic datasets is risky, because the data may fit one of the algorithms better than the others.*" The only reliable approach is to use real data from actual users of the proposed recommender system; however, it can take a long time to collect enough rating data and social data. It has been found that it takes almost two years to overcome the cold start problem [106].

Because of these reasons, we decided to postpone the research recommendation using multi-criteria ratings with taking into account the social relationships between users for the future, and this will be done by designing and deploying a system that could be tested with real users. Due to scope and time constraints on the PhD thesis, we dedicate the rest of the thesis on exploiting social relations of users in social bookmarking websites for research papers to improve not only the performance of the recommendation (i.e. prediction accuracy) but also the non-performance represented by the user coverage of recommendation and the diversity of the recommended list.

3.4 Summary and Conclusions

In this chapter, the focus group study was discussed which aimed to elicit the users' views on rating the papers using different rating criteria evaluating different paper quality aspects. In addition, we discussed with the participants issues related to user control of the rating criteria and the user interface design for research paper RSs. Another study also conducted to confirm the findings from the focus group. The study focused on the ranking of the rating criteria that we can use to build a new paper recommender system, and also to gather the user opinion regarding the user's control on the importance weight of the rating criteria. The results can be summarized as follows:

- Participants thought that using different rating criteria is very useful to find more personalized recommendations.
- Participants showed the strongest support for the following three criteria: clarity, technical clarity and relevance.
- Participants liked the idea of controlling the weight of the different rating criteria and observing the resulting change in the list of recommendations. User control seems to contribute in the level of trust in the recommendations and in the RS.
- Trusting the RS can be increased if the users can provide the RS with feedback to tell the RS whether it satisfies their needs (i.e., making the system scrutable [124]).
- It is important to find a way to judge fairly the recently published papers rather than relying on the number of citations as a criterion for paper quality.
- It is important to make the RS more dynamic by giving some control to the user such as the ability to change the weight of each rating criterion and choosing the ranking order of the recommended papers.
- Paying more attention to the presentation style of the recommendation as well as the user interface design can increase user satisfaction and trust in the RS.
- Explaining to users how the recommendation list is chosen is important in their accepting the recommendation. The participants of this study only guessed about the recommendations that they saw in some of the paper recommendations, such as Academia.edu and ResearchGate (see the list of their hypotheses in section 3.1.2).

- The participants in both studies agreed on the same rating criteria ranking and also agreed on enabling the users to give different weights for the rating criteria to increase the personalization of the recommendation.

As we noticed from the quantitative study, participants have some differences mostly in the ranking for “clarity” and “technical clarity”, and in the case for master student, who gave more weight to the “value added by reading the paper”, which supports the idea of giving the user the ability to change the weight of these criteria. Since users have different objectives and different needs, changing the weight of these criteria can help them to cope with these changes.

We can conclude that, in in our future, when building a research paper recommender system with multi-criteria rating, the following criteria will be used (as summarized in Table 3.5):

- Relevance
- Value added by reading this paper
- Technical clarity

In the next chapters, we present our work exploring how implicit social networks can be used in the recommendation of scholarly papers to improve both performance and non-performance metrics.

CHAPTER 4:
IMPLICIT SOCIAL NETWORKS AND INTEREST SIMILARITY

In section 2.3, we studied extensively the state of the art of the social recommendations, and we found that the social recommendation is beneficial in increasing the prediction accuracy, the user satisfaction and most importantly in alleviating the effect of the cold start problem of CF. In addition, as shown in sections 2.3.4 and 2.4.5, all the work in the area of recommending research papers using the social relations between users is based on explicit social relations. However, it is known that recommendations are based on explicit social relations suffer from low user coverage. Therefore, we explore the effect of social relations between users that are built based on the collected user interaction data and reflecting interest similarities between users. These relations are not initiated by users, but calculated by machine based on the user bookmarking behaviour. We want to test if exploiting such relations can improve some *performance* and *nonperformance* measures of recommendations. We propose three implicit social networks so that users have an equal opportunity to be connected to any of these three. We mainly focus on the user's bookmarks and the user-generated data (i.e. tags) that are available in the social bookmarking tools.

In this chapter, we explore the similarities in the users' interests and their inferred relations through their behaviours on social bookmarking Web sites. We construct three social networks that connect users implicitly based on their paper bookmarking behaviour. Then we compare the user similarity in each network separately based on the distance of the relation between two given users. Afterwards, we compare the three proposed implicit social networks with each other. The objective is to harvest the data available in the social bookmarking tools to find more users similar to the target user. An advantage to our approach is that users do not need to put in more effort by providing the system with new input. In [125], the authors show that the overlap between the connected users and the most similar users is only 9.23%, which means that users are mostly not connected to their similar users. In our work, we try to reduce the gap between the connected users and the most similar users groups. This could be done by connecting users using implicit social networks considering their interest similarity. Proving the usefulness of these networks as information sources is beneficial to improving hybrid paper recommender algorithms that use the proposed implicit social networks. The rest of this section is organized as follows: we describe the three proposed implicit social networks and the dataset and similarity measures that are used in the experiments. After that, we explain the two experiments to evaluate what depth and type of network is most useful in discovering users with similar interests.

We build three implicit social networks based on the bookmarking behaviour of users from the data we collected from CiteULike.org which is described in section 4.5. The next sections describe the proposed networks.

4.1 Network 1: Readership ISN

Here, we build a social network based on the readership relations between users who are members of CiteULike. We connect users to the authors of the papers that they have bookmarked in their libraries. We assume that if users bookmark specific papers, the bookmarkers and the authors of the papers have interest overlap; this overlap increases with an increase in the number of papers bookmarked from the same author. The relation could be unidirectional or reciprocal. The relation is unidirectional if only one of the users in this relation has bookmarked the other user's publication. The relation is reciprocal if both users have bookmarked each other's publications. Figure 4-1 shows the relations in this network depicted as black arrows. For example, the relation between user 3 and user 5 is reciprocal, while the relation between user 3 and user 1 is unidirectional; user 3 is the bookmarker of the paper and user 1 is the author of the paper. To avoid making the graph complicated, we assume that the strength in the reciprocal relations is the same for both directions. However, one of the users could bookmark more or fewer papers from the other user who is involved in the reciprocal relation. The numbers on the arrows represent the strength of the relations. For example, the strength of the relation between user 3 and user 1 is 5, which means there are 5 bookmarked papers in user 3's library that were authored by user 1.

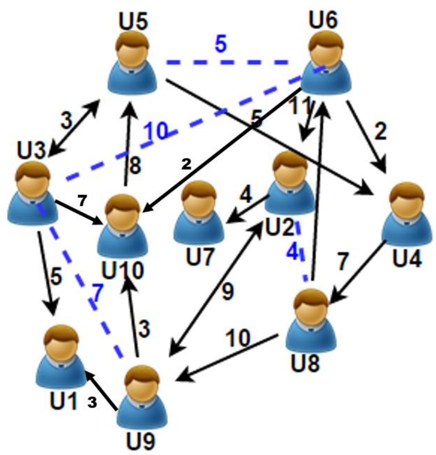


Figure 4-1: Sample of relations in implicit networks

As a first step to constructing this network, we select the authors of the papers represented in the dataset. The users who bookmarked these papers are selected as step 2. Step 3 involves collecting the publication list and the bookmarked list for all users, along with the data of the paper, such as the date of adding the paper to the user's library, the user's tags associated with each paper and the metadata of the papers. The relations are constructed between the members of the two user sets obtained in steps 1 and 2, and their strengths are calculated as the number of papers by a specific author that the bookmarker user has bookmarked. The data is organized into three columns (authorID, bookmarkerID, strength). Because not every author of a paper in the dataset has a CiteULike account, nor does every author in the dataset have his or her publications bookmarked by users in the dataset, 4,909 users are involved in unidirectional readership ISN and 209 is the number of users involved in reciprocal readership ISN.

4.2 Network 2: Co-readership ISN

In this network, users who bookmark (and presumably read) papers written *by the same authors* are connected. If user 1 and user 2 have both bookmarked papers written by user 3, then user 1 and user 2 are connected using the co-readership implicit social network.

This network structure is useful for users who do not have publications yet; therefore, they cannot have relations in network 1. The assumption is that users who bookmark the paper(s) that are written by same authors have similar interests. The strength of the relationship is measured by the number of authors that overlap in their libraries. The higher the strength, the more similar the users. Figure 1 shows an example of the relationships in this network in blue. For example, user 5 and user 6 are connected because they both bookmarked papers written by the same authors; here the number of overlapping author names is five. We show only part of the graph, and it includes only one of those five authors (user 4).

To construct this network, as step 1, the author list for each paper is extracted from the paper's metadata, each author name is assigned an ID and data is stored so that each authorID is associated with the corresponding paperID. As step 2, for each user, we find the authors list associated with his or her bookmarked papers and store each user associated with the authorIDs of his or her bookmarked papers. After that, as step 3, we look for users associated with the same authorIDs and create the relations. The data is organized into three columns (user1, user2, strength), where the strength is calculated as the count of common authors the two users are associated with. The number of users who have connections in this network is 11,508.

4.3 Network 3: Tag-based ISN

In this network, users are connected if they use the same tags to annotate their bookmarked papers. However, we do not check whether users use the same tags to annotate the same papers. We consider the tag similarity between the entire tag cloud associated with each user. We assume that the more similar tags the users have, the higher the interest similarity. While the previous two networks were based on the metadata of the papers, this network is based on user-generated data. The relations in this network also have strengths. The strength of the relation between two users is measured by the number of tags shared between them.

To build this network, the tags used to annotate the papers are aggregated for each user. The data is preprocessed to make the tags comparable. We follow the method described in [126] to preprocess the tags. All tags are preprocessed by converting them to lowercase, removing the stop words, and then using the porter stemmer tool to remove any additional letters added to the root word to eliminate the effect of the word variation (e.g., the word "social" could have different variations, such as "socialize", "socialization" and "socializing").

After that, the relations between the users are constructed. A connection between two users is initiated if both users use the same tag to annotate any papers in their libraries. The data is also organized into three columns (user1, user2, strength), where user1 and user2 represent the ID numbers for the respective users and strength is calculated as the number of tags shared between the two users. The assumption is the more tags shared, the stronger the relationship. There are

11,283 users included in this network. This is because there are users who did not use any tags to annotate their papers.

4.4 Propagation of Relations within Social Network

Social networks allow the propagation of influence, or they indicate affinity among people along paths in the network. This “propagation” property allows conclusions to be drawn about similarities among three people connected with each other, and it can help deal with the cold start and data sparsity problems in recommender systems (e.g., [73]). This is beneficial in the recommendation stage to help find people who are not connected directly but still have relevant interests. For this reason, we expand each of the three networks discussed above in depth to explore the similarity between two distantly connected users. Three depths of social relations are considered. First, if the users are directly connected using the defined relations in each network, the relationship is considered a direct relationship. For example, in the readership network (network 1), user 1 and user 2 are directly connected if user 1 bookmarks a paper(s) authored by user 2; see Figure 2. In the co-readership network, user 1 and user 2 are directly connected if both bookmark a paper(s) that is authored by a specific author name. The indirect relations connect users if they are connected by an intermediate user; the distance between those users could be one user (called one hop) or two intermediate users (called two hops). As shown in Figure 4-2, the relations are direct between (U1,U2), (U2,U3) and (U3,U4); indirect with one hop between (U1,U3); and indirect with two hops between (U1,U4).

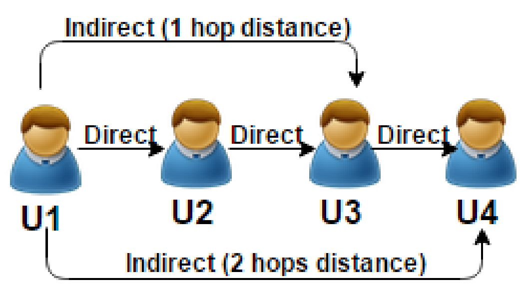


Figure 4-2: Different social relation distances in respect to U1

4.5 Dataset and Similarity Measures

4.5.1 The Dataset

The data used for this study was collected from the CiteULike.org social bookmarking Web site. CiteULike is used to save and organize the researcher’s interesting resources (e.g., books, journal articles, conference papers, Web sites). CiteULike allows social features such as connecting users, watching users (like following on Twitter) and sharing scientific references among researchers. Users of CiteULike can import scientific reference data from other resources such as PubMed and can assign tags to the bookmarked references for future easy access. In addition, users can be aware of who has the same references they do.

CiteULike offers a dataset for researchers that contains the user identification number, paper identification number, date and time the paper was bookmarked by the user and tags that the user has used to annotate the paper. For our study, however, we also needed the paper content–related data contained in the metadata for the published and bookmarked papers (i.e., title, authors, abstracts and publication years). For this reason, we collected the needed data by crawling the CiteULike Web site. The data was gathered using the snowball method. We first randomly chose 482 recently active users as initial users and then collected their publications and bookmark data. Then we branched to collect the users' data for the users who had bookmarked their publications or who had bookmarked the same papers as the initial users. The descriptive statistics for the dataset collected between December 2014 and February 2015 are shown in Table 4.1.

Table 4-1: Descriptive analysis of the dataset

Initial number of users	482
Total number of users	13,189
Total number of distinct papers	1,043,675
Total number of publications	19,774
Total number of bookmarks	1,323,065
Total number of tags	3,086,565
Average number of bookmarked papers per user	98.79 (σ: 91.70407)
Average number of bookmarkers per paper	1.251 (σ: 1.647203)
Average number of tags per user	3.81
Average number of publications per user	1.523846 (σ:6.841809)
Average number of authors per user	389.2183 (σ:527.441)
Number of users who have publications	2,508
Number of users in unidirectional relations	4,909
Number of users in reciprocal relations	209
Total number of total unidirectional relations	9,248
Total number of reciprocal relations	141

4.5.2 Similarity measures for user interests

Different similarity measures are needed because there is no empirical evidence that indicate the best similarity measure for the unary data (i.e. bookmarked-based user preferences) one perfect measure that captures the complexity of user interests. The existing literature metrics differ with respect to their scope and complexity of computation. We discuss below four of these that are appropriate in our case. Cosine similarity measures the similarity of the paper text, and the rest of the measures are mainly used to compare the similarity for binary ratings.

4.5.2.1 Number of Co-bookmarked Papers

The most obvious and simple measure of similarity between the interests of two users can be defined as the overlap between the bookmark libraries of the users. We can consider the act of a user adding the paper to her library a binary user rating of the paper; there are many ways to compute similarity of users based on their ratings, the simplest one being the count of similar ratings given to the same items. The more papers in common between two users, the more similar they are.

4.5.2.2 Jaccard Coefficient

Even though finding the similarity between two given users through counting the number of shared bookmarked papers is easy and straightforward, it does not take into account the size of the library of those users. An overlap of five papers between two very large libraries may not indicate a great similarity of the interests of the two users, whereas if the sizes of their libraries are small (e.g., 10 papers in total), this would be an indication of high similarity.

To make the similarity relative to the size of the users' libraries, it is beneficial to use relative similarity measures such as the Jaccard coefficient, which is a ratio between the number of co-bookmarked papers to the union of the bookmarked papers between the two given users. With the assumption that we have two users, U1 and U2, and the bookmarked papers of U1 and U2 libraries are given by L_{U1} and L_{U2}, respectively, the Jaccard coefficient (J) between U1 and U2 can be calculated by the equation (Eq. (4.1)):

$$J(U_1, U_2) = \frac{|L_{U_1} \cap L_{U_2}|}{|L_{U_1} \cup L_{U_2}|} \tag{4.1}$$

The value of the Jaccard coefficient is a continuous value between 0 and 1. The closer the value is to 1, the more similar the users are.

4.5.2.3 *Log-likelihood Similarity*

While the number of co-bookmarked papers and the Jaccard coefficient consider only the intersection in bookmarked papers between two users, the log-likelihood considers if both users did not bookmark certain papers and the cases when one user bookmarks papers that are not bookmarked by the other user [127]. The assumption is that if both users did not bookmark some papers, they have similar preferences in what papers they do not bookmark. The log-likelihood ratio measures how unlikely the similarity is due to chance and whether the similarity between two given users represents genuine similarity. If we denote n_1 and n_2 the number of papers bookmarked by $user_1$ and $user_2$, respectively; $n_{1,2}$ the number of papers that are bookmarked by both user 1 and user 2; and $n_{\sim 1\sim 2}$ the number of papers that are bookmarked neither by user 1 nor by user 2, the log-likelihood ratio (LLR) between user 1 and user 2 is calculated by equation Eq. (4.2):

$$\text{LLR} = 2(\text{H}(n) - \text{H}(\text{rowSums}) - \text{H}(\text{colSums})) \tag{4.2}$$

where H is Shannon's entropy, computed as follows:

$$\sum \frac{n_{ij}}{\sum n} log \frac{n_{ij}}{\sum n} \tag{4.3}$$

where i,j are the users numbered (1,2),

$$\text{H}(\text{rowSums}) = \text{H}\left(n_{1,2}, n_2\right) + \text{H}(n_2, n_{\sim 1\sim 2}) \tag{4.4}$$

$$\text{H}(\text{colSums}) = \text{H}\left(n_{1,2}, n_1\right) + \text{H}(n_1, n_{\sim 1\sim 2}), \tag{4.5}$$

$$\text{H}(\text{n}) = \text{H}(n_{1,2}, n_1, n_2, n_{\sim 1\sim 2}) \tag{4.6}$$

4.5.2.4 *Cosine similarity for textual metadata*

Even if the overlap between two users' bookmarked paper collections is small or nonexistent, the users may still have similar interests. However, a text analysis of the users' bookmarked papers may suggest similar terms that indicate semantic similarity.

To measure the similarity between two users according to the text of the bookmarked papers, the vector-space model [128] is used to represent the user's preference model. All terms that appear in the titles and abstracts of the bookmarked papers are aggregated as a bag of words. We used the title and the abstract of the papers because they are publicly available and because it has been found that terms from the title are three times stronger than the body text and from the abstract twice stronger than the body text [129]. For better comparison, it is beneficial to apply some text analysis processing to the bag of words as we did for the tags discussed when we described network 3. First, all the terms were converted to lowercase letters and then the stop words were removed. After that, the applied Porter stemmer truncates any additional letters and converts the word to its root. The processed bag of words is then converted to a keyword vector, which consists of tuples (term, TF/IUF), where TF/IUF means the ratio of the term frequency over the inverse user frequency. Following the modified version of TF/IDF in [10], we substitute the document frequency in TF/IDF (term frequency/inverse document frequency) with user frequency, that is, how many users have bookmarked papers that contain the term in their title or abstract. After the vectors for all users are built, the cosine similarity can be applied on the two user vectors that we want to calculate the similarity for. The cosine similarity is the cross product of both vectors of users. The more overlap in terms between two users, the more they are similar.

4.6 Experiments, Results and Discussion

The objective of this study was to compare the similarity between connected users in the three proposed social networks and to measure the similarity between connected users involved in direct and indirect relationships. The experiments, results and discussion of the results are given in the following sections.

4.6.1 Comparison of Interest Similarity among the Three Networks

To explore whether there is any difference in user interest similarity among the three implicit social networks and to answer the following research question:

> **RQ2:** Comparing three implicit social networks, readership (which consists of one of two types of relationships: reciprocal or unidirectional), co-readership, and tag-based, which one connects the most similar users?

We conduct an experiment to test the following null hypothesis to test the interest similarity values among different ISNs and for each social distance:

$H0_{4.1}$: There is no statistical difference between the means of interest similarities of directly connected users and indirectly connected users using one hop and two hops of distance among three implicit social networks.

We conduct a two-way ANOVA test to compare the means of similarities between users in the three social networks considering the social relation distance—direct, one hop indirect and two hops indirect—among the three networks. The results show statistically significant differences ($p < 0.01$) between the means of the interest similarities for all measures. Hence, the null hypothesis $H0_{4.1}$ is rejected.

Table 4-2: Pairwise comparisons of two-way ANOVA test to compare the three networks

	LogLikelihood	No. of co-bookmarked papers	Jaccard coefficient	Cosine similarity
Direct	1>2>3>4	1>2>3>4	1>2>3>4	1>2>4>3
	F*=27003 *	***F*=6258.502*** *	***F*=6492.004*** *	***F*=979.492*** *
Indirect(1 hop)	1>3>2>4	1>3>2>4	1>3>2>4	1>2>4>3
	F*=13588.466 *	***F*=3730.612*** *	***F*=3780.111*** *	***F*=1828.934*** *
Indirect(2 hops)	1>3>2>4	1>3>2>4	1>3>2>4	1>2>4>3
	F*=10078.120 *	***F*=2646.925*** *	***F*=2529.388*** *	***F*=4642.810*** *

1= Reciprocal readership implicit social network
2= Unidirectional readership implicit social network
3= Co-readership implicit social network
4- Tag-based implicit social network
* **Significant at p<0.01**

Moreover, the post hoc pairwise comparisons Scheffé test [130], which is suitable for comparing results of different group sizes, show that all the results are consistent (see Table 4.2): the reciprocal readership network has the highest similarity between connected users, the tag-based network has the lowest similarity for all distances. However, in the direct relations, the unidirectional readership network is slightly higher than the co-readership network for direct relations, with an average difference of 0.33, and the co-readership network is slightly higher than the unidirectional readership network for indirect 1-hop (average difference = 0.16) and 2-hop (average difference = 0.41) relations. This is true for all the similarity measures except the cosine similarity. When the text-based cosine similarity measure is used, the tag-based network performs slightly better than the co-readership social network, with a mean difference of 0.02.

The results can be interpreted as users being more similar to the authors of the papers that they bookmarked in their libraries, and the similarity increases if the relation is reciprocal. Furthermore, even when users do not bookmark papers written by the same authors, they construct a community of users who share similar topics with users who are directly and indirectly connected. It can also be inferred that the user-generated data (i.e., tags) does not do better than the metadata that is used to construct the other implicit social networks. This might happen because users cannot use the best representative tags for the bookmarked papers due to users in the dataset have very few tags. As shown in Table 4-1, the average number of tags per user is only 3.81.

4.6.2 Impact of the Relation Distance on Users' Interest Similarity

We first explored the effect of the distance of the relation between two connected users on their interest similarity in each of the three implicit social networks defined above. The research question that we aim to answer is:

> **RQ3:** In each of the three proposed ISNs, how does the relationship distance between two connected users affect the users' interest similarities?

As a baseline hypothesis we used:

$H0_{4.2}$: There is no statistical difference between the mean of interest similarities of directly connected users and indirectly connected users using distances of one hop and two hops for each of the three implicit social networks.

To test the above hypothesis, we conduct a one-way ANOVA test. The results are shown in Table 4-3. For all the interest similarities measures, there is a statistically significant difference in the similarities depending on the social distance of user pairs for all three implicit social networks. Thus, the null hypothesis $H0_{4.2}$ is rejected.

Table 4-3: Pairwise comparisons of one-way ANOVA test to compare the three networks based on the relation distance

	Distance	Loglikelihood similarity	No. of co-bookmarked papers	Jaccard coefficient similarity	Cosine similarity
Reciprocal readership ISN	Direct	0.81260	9.579	0.03860	0.34485
	Indirect (1 hop)	0.44246	3.738	0.00992	0.33059
	Indirect (2 hops)	0.47059	1.125	0.00278	0.26786
	F value	**27.435***	**13.115***	**8.615***	**6.847***
Unidirectional readership ISN	Direct	0.84535	1.180	0.00954	0.2973
	Indirect (1 hop)	0.14673	0.253	0.00130	0.2881
	Indirect (2 hops)	0.12986	0.198	0.00209	0.2948
	F value	**22999.746***	**885.378***	**340.166***	**30.778***
Co-readership ISN	Direct	0.31183	0.736	0.00493	0.31075
	Indirect (1 hop)	0.26066	0.601	0.00388	0.30779
	Indirect (2 hops)	0.22334	0.499	0.00318	0.30626
	F value	**2669.369***	**885.738***	**1194.787***	**72.640***
Tag-based ISN	Direct	0.09320	0.251	0.00106	0.31138
	Indirect (1 hop)	0.06359	0.147	0.00082	0.31136
	Indirect (2 hops)	0.05380	0.115	0.00047	0.30320
	F value	**1675.685***	**826.782***	**734.432***	**337.251***

* **Significant at p<0.01**

We also apply post hoc pairwise comparisons Scheffé test, which show that across all networks and for all similarity measures, users who are involved in direct relationships have the highest similarity and the similarity decreases with the increase of the social distance. This means users connected in each of the three networks share similar interests and collect similar papers in all the tested social distances. Therefore, the relations in all the proposed implicit social networks are transitive.

According to Tang and Liu, users in the same group share similar interests (or preferences), and they establish weak connections if they are connected distantly [131]. This is exactly the results that we got using our implicit social networks. We believe that social recommending algorithms would benefit most by exploiting the direct relations between users to recommend papers from the most similar users to the target user. Granovetter suggests that novel information flows between people who have weak tie connections [132]. Sugiyama & Kan found that serendipitous recommendations for the target user could be obtained from her dissimilar users [133]. Hence, the indirect relations can be good resources to enrich the user's library with diverse papers and help make connections with new people.

4.7 Comparison between Implicit Social Networks and Explicit Social Networks

In section (4.6), the similarities between users who are involved in each of the proposed implicit SNs are compared in two ways. First, for each SN, similarities of users who are connected using different connection distances are compared. Then, the similarities between different users in different implicit SNs are compared. However, comparing the similarities between users in implicit SNs with explicit SNs is necessary to discover if the proposed implicit SNs is worth discovering. In this section, we compare the similarities in the implicit SNs with two explicit SNS: co-authorship SN, and friendship SN between users, which is constructed by the invitation/acceptance between users who are involved in the social relation aiming to answer the following research question:

> **RQ4:** Does the interest similarity between users who are implicitly socially connected comparable to the one between users who are explicitly connected?

4.7.1 Co-authorship Explicit Social Network

The co-authorship relations between two users happen when they collaborate in writing and publishing a research paper(s). When two users collaborate in publishing papers, this means that they share similar interest and have strong relationship. The co-authorship SN also has a strength represented by how many papers the pair of users has co-authored.

In CiteULike, users can declare that the paper is one of their publications. From the list of publications for each user, the set of co-authors are collected and organized. The data for co-authors network is presented in Table 4-4. Only 247 users out of 13,189 in the whole dataset are involved in co-authorship relations. It is important to mention, in this research, we only consider co-authors within the CiteULike community. However, not every user enter their publications. Most of the studies that are done using co-authorship consider the author names extracted from the paper text, so that they could gather more relations.

Table 4-4: The descriptive statistics for co-authorship network

Number of co-authors	247
Total number of social relations	167
Average number of social relations per user	1.274809
Number of the co-authors' publications	4,181
Average number of publications per user	16.9271
Number of their bookmarks	43,134
Average number of bookmarks per user	174.6316

Co-authorship relation is considered explicit social relation since both users involved in the relation know each other, are aware of the relationship, and have engaged in it voluntarily. We want to compare the co-authorship network to the proposed implicit social networks to see how these networks perform in identifying similar users. Can the implicit social networks identify the users connected through the explicit social network? If the answer is yes, this would indicate that the implicit networks are able to discover users' preferences.

4.7.1.1 Which Implicit Social Network Discovers the Co-authorship Relations?

The social relations in the co-authorship networks were compared with the social relations in each network and we found that 98 out of 167 social relations (58.68 percent) in the co-authorship was discovered by the bi-directional relations in network1: readership network; 60 social relations (35.92 percent) were discovered by the unidirectional relations in network1; and 152 social relations (91.01 percent) were discovered by tag-based social network and none were discovered by co-readership network. This is because users in this network were selected from those who have not authored publications (i.e., they are bookmarkers only) and users who had authored publications but have not been bookmarked by anyone (i.e. not in readership network).

These results are quite good, especially for the tag-based network; authors showed that they have large similarities in their tag clouds so that tag-based social network discovered most of the co-authors. The results of readership network could be better if all authors bookmarked their own publications since CiteULike allows that. However, co-authorship network only covers 1.87 percent of the users in the dataset.

4.7.1.2 Comparing the Interest Similarity between Users in Co-authorship with the Interest Similarity in the Implicit Social Networks

We compared the interest similarity between users in co-authorship network with the interest similarity of users in the proposed implicit social networks. The null hypothesis is:

$H0_{4.3}$: There is no statistical difference between the means of the interest similarity between users connected in co-authorship SN and the means of the interest similarity between users connected by different implicit SNs

One-way ANOVA test was conducted which showed that there was statistical difference between the mean values of the similarity measures of the different networks. Table 4-5 shows the means values and the ANOVA test results and the post hoc test results. So, the null hypothesis $H0_{4.3}$ is rejected.

Table 4-5: Pairwise comparison between explicit social networks and implicit social network

	Network	LogLikelihood similarity	No of co-bookmarks	Jaccard coefficient	Cosine text vectors similarity
1	Reciprocal readership ISN	0.8126	9.58	0.0386	0.4231
2	Unidirectional readership ISN	0.8454	1.18	0.0095	0.3489
3	Co-readership ISN	0.3119	0.74	0.0049	0.3027
4	Tag-based ISN	0.0932	0.25	0.0011	0.3223
5	Co-authorship network (explicit SN)	0.5920	6.75	0.0254	0.4136
6	Friendship networks (Explicit SN)	0.4401	2.90	0.0116	0.3987
One-way ANOVA test results		*F*-value= **14442.670***	*F*-value= **4193.989***	*F*-value= **3520.992***	*F*-value= **898.304***
Scheffé post hoc test results		1>2>5>6>3>4	1>5>6>2>3>4	1>5>6>2>3>4	1>5>6>2>4>3

*** Significant at p<0.01**

Then the Scheffé post hoc test showed that co-authors had a lower interest similarity than the reciprocal readership implicit SN, but higher than the other implicit SNs. This was true for all similarity measures except the Loglikelihood in which connected users in reciprocal and unidirectional readership implicit SN had higher similarity than connected users in co-authorship network. All the results were significant at p<0.01 except for the cosine similarity where the reciprocal readership mean value was higher than the co-authorship mean value but the difference was not significant. This means that the interest similarity between users in both networks is the same. The cosine similarity measure that is based on the text vectors containing the title and abstract of bookmarked papers show that users collect similar papers, but not the exact same papers. Even though the results show that the co-authorship network is able to connect users with higher interest similarity, only users with publications that are co-authored with other users can be part of this network, which means only1.873 percent of the users.

4.7.2 Friendship Explicit Social Network

The social relation in the friendship explicit social network is undirected relation that happens between two users when one user invite the second user to add her to her connection list (i.e. contact list), and the second user accepts the invitation. In CiteULike, the list of friends are called “connections”. Because the term friendship is more known for this kind of relationship, so hereafter, the term “friendship” will be used to denote the user whom the target user is connected to explicitly. In order to compare the proposed implicit social networks to the friendship network, the data of the friends of the target users were collected from CiteULike. For each friend, the publication and bookmark lists were extracted. Table 4-6 shows the data collected for the friends. As shown only 2375 out of 13,189 users in our dataset have social relations using the friendship network. This means users in social bookmarking websites are focusing on bookmarking papers more than the social aspects of the websites.

Table 4-6: Descriptive analysis of friendship network

Number of users who have friends	2375
Number of publications of the friendship network	10,257
Number of bookmarks	360,715
Average number of bookmarks per friend	99.152
Number of relations in friendship networks	6,171
Number of distinct friends	3,638
Average number of friends per user	0.311

The same approach of comparing users in different implicit social networks and co-authorship SN was also used here. Using different similarity measures, the interest similarity of users involved in friendship social relations were compared with the ones for users who were involved in direct social relations using the proposed implicit social network.

The null hypothesis was:

$H0_{4.4}$: There is no statistical difference between the means of the friendship SN and different implicit SNs

One-way ANOVA test was conducted which shows that there was statistical difference between the mean values of the similarity measures of different networks, please refer to Table 4-5. So, the null hypothesis $H0_{4.4}$ was rejected. Then the Scheffé post hoc test was applied which show that the friendship SN have lower interest similarity than the reciprocal readership implicit SN but higher than the other implicit SNs. This was true for all similarity measures except the Loglikelihood in which the reciprocal and unidirectional readership implicit SN did better than the friendship network. All the results were significant at $p<0.01$. Even though the results showed that the friendship network was able to connect users with higher interest similarity, only users with explicit social relations can be part of this network, which comprise only 18% of all users.

4.8 Summary

In this chapter, we tested the feasibility of discovering interest similarity between users in three proposed implicit social networks that are built based on the users' bookmarking behaviour in social bookmarking tools. For each network, we compared the interest similarities between users who were involved in direct relations, as well as indirect (one-hop and two-hops) distance relations. We found that there was a statistically significant difference in users' similarities with respect to the relation distance for a given network. Next, the interest similarities among the three different social networks were compared, and the differences were also found to be statistically significant. In addition, we compared the interest similarities of users in the proposed implicit social networks with the users' interest similarities in two explicit social networks.

The big advantage of using implicit social networks is that they do not need explicit actions from the user such as watching/following or connecting to other users. Instead, only the data that the user gathers in their private behaviour through bookmarking papers (publicly available data) is used to create the networks. It is also not necessary for users to have publications; all users are treated equally in constructing the proposed social networks so that junior and senior researchers can benefit from the proposed approach.

CHAPTER 5: EXPLOITING THE IMPLICIT SOCIAL NETWORKS IN DIFFERENT RECOMMENDING APPROACHES

In Chapter 4, we constructed three social networks based on implicit rating of research papers. We calculated and compared the interest similarity between users involved in each network. However, we did not evaluate or compare the recommendation performance of any of the proposed ideas. We aim to evaluate the recommendation produced using implicit social networks as sources of recommendation. Unlike the majority of previous studies which focus on increasing the prediction accuracy only, we study the effect of using implicit social relations on *prediction accuracy* and *user coverage*, and the *trade-off* between them.

In this chapter, we explain the recommendation approaches and the evaluation measures that we use in the experiments. Then we define the experiments, objectives, research questions, and hypotheses and present the results.

5.1 Recommendation Approaches

We compared various existing recommendation approaches to explore if using different sources of information from implicit SNs will affect the prediction accuracy of the approaches. We compared the following recommendation approaches:

5.1.1 Social Recommender (SR)

It is a simple approach to incorporate the social information into the user-based CF. Social recommender was proposed by [57]. It simply replaces the anonymous nearest neighbors in the user-based CF with the target user's social friends in the social network. To apply the social recommender to the proposed ISNs, we found the social friends of each user and used the data from those friends in the same way that anonymous peers in CF are used - by picking the top N peers and using their bookmarked papers to find candidate papers to recommend to the user.

However, in the social recommender we replaced the similarity between users that is used in the prediction of the target user's rating for unseen items with the weighted strength between users Ui and Uj $\mathrm{WStr}_{\mathrm{U}_{i,j}}$ calculated as:

$$\mathrm{WStr}_{\mathrm{U}_{i,j}} = \frac{\mathrm{Str}_{\mathrm{U}_{i,j}}}{\mathrm{TotalStr}_{\mathrm{U}_i}} \tag{5.1}$$

Where $\mathrm{Str}_{\mathrm{U}_{i,j}}$ is the strength of the relation between Ui and Uj and $\mathrm{TotalStr}_{\mathrm{U}_i}$ is the sum of all strength values between Ui and all of other users who are connected to her.

5.1.2 Combined Recommender (CR)

The combined recommender is a hybrid approach that integrates neighbors from conventional user-based CF and the target user's social friends to construct a new nearest neighborhood set for the target user [57]. We then used the data from users in the new combined neighbors in the recommendation following the same way as in the social recommender.

5.1.3 Amplified Recommender (AR)

The amplified recommender is a hybrid approach that amplifies the social friends' preferences in CF nearest neighbors [58], which amplifies the friends' preferences in CF with nearest neighbours. First the nearest neighbourhood peers are identified by CF top-N technique. Then if the user's social friends are also in the top-N neighbours, an amplifying approach is used to give the preferences from those social friends more weight in the recommendation process. The amplifying function that we will use is the one used in [58] which is given by Equation. (5.2):

$$\mathrm{Min}\ (\mathrm{S}_{\mathrm{U_iU_j}} \times (1 + \frac{\mathrm{N}_{\mathrm{U_iU_j}}}{\mathrm{N}_{\mathrm{all,U_j}}})\ ,1) \tag{5.2}$$

where Uj is the target user, Ui is one of the Uj's social friends, $\mathrm{S}_{\mathrm{U_iU_j}}$ is the similarity between Ui and Uj which is calculated by CF approach using the papers that are co-bookmarked by both users, $\mathrm{N}_{\mathrm{U_iU_j}}$ is the number of interaction between the target Uj and the user's social friend Ui, and $\mathrm{N}_{\mathrm{all,U_j}}$ is the total number of interactions between the target Uj and all of the user's social friends.

Because the similarity value cannot be greater than 1, we chose a minimum value between 1 and the amplified value. The interactions between the target user and one of the user's social friends were based on the type of ISN on which we were trying to apply the approach. For example, if we use the co-readership ISN, the number of interactions equals the number of authors that both users have in common (i.e., the number of authors one or more of whose papers both users bookmark).

5.2 Evaluation Measures

Evaluation of the recommendations depends on the research questions that need to be addressed as well as the methods of the evaluation. Examples of research questions that can be asked are: Are the recommended items relevant to the user? Does the recommender system suggest diverse items? Novel items? Does the recommender system scale well? Does the recommender system cover most of the users? The evaluation of the recommender system could be done using offline or online experiments. Offline experiments are mainly used when the focus is to evaluate the performance of algorithms that do not need the user's intervention, while online experiments focus on the user experience such as the user satisfaction and acceptance of the system. In the rest of this dissertation, the focus is to test the performance of the recommendations produced using different ISNs that are proposed in Chapter 4. All the evaluations are done offline using users' bookmarking histories. There are many evaluation measures for unary ratings (presence of the paper in the user's library). Some are for measuring the performance of algorithms while the other are nonperformance measures.

5.2.1 Performance Measures

In order to evaluate the relevancy of recommended items using the aforementioned implicit SNs as resources of recommendation, information retrieval-based evaluation methods usually are used such as precision, recall and F-measure. Offline experiments are usually conducted using a method called N-fold cross-validation. It is a random selection technique in which one fraction of the user's bookmarks of size (1/N) is selected as a testing set and the remaining (N-1)/N fractions of the user's bookmarks are used to train the algorithm's model. This process is repeated N times, each time with different test and training sets. Then the accuracy of the prediction is calculated. In this dissertation, fivefold cross validation is used for all offline experiments where 20% of the user's bookmarks are used as testing data and 80% are used as training data. This process is repeated five times, each time with different test and training sets. Then the accuracy of the prediction is calculated. In Table 5-1, different recommendation output is shown with respect to the items in the user's library.

Table 5-1: Possible output of recommender system [13]

		Recommended	Not Recommended
Liked by the user!	Yes	True positive (TP)	False negatives (FN)
	No	False positives (FP)	True negative (TN)

Precision is defined as the ratio of the true positives in the recommended list to the total number of items in the recommended list, and the recall is the ratio of the true positives to the all items in the test set. Precision (P) and recall (R) are given by the following equations.

$$P = \frac{TP}{TP+FP} \tag{5.3}$$

$$R= \frac{TP}{TP+FN} \tag{5.4}$$

Precision measures the number of papers recommended and liked by the user to the total number of recommended papers, while the recall measures the ratio of the papers that are recommended to the user to the papers in the test data.

It is always assumed that the items with higher ranks in the recommended list of items are more important than items with lower ranks. When precision and recall evaluation measures are used, three ranks are considered: top 2, top 5 and top 10. Then we compare the results among these ranks. Precision@N and Recall@N are calculated with respect to the rank. For example, if Precision@10 is used, we calculate the ratio of true recommended items to the top 10 recommended items, and the Recall@10 is the ratio of the number of true recommended items in the top 10 recommended items to the test set.

When N, the number of recommended items, increases, a trade-off between precision and recall measures is observed. When N increases, the precision value starts to decrease, while the recall starts to increase. To reduce the effect of the change of the precision and recall by increasing the N value, the F1-measure can be used to produce evaluation results that are more universally comparable. F1 can be calculated using the following equation:

$$F1@N=\frac{2.P@N.R@N}{P@N+R@N} \tag{5.5}$$

5.2.2 Nonperformance Measures

Measures that are discussed in section 5.2.1 are used to measure the performance of prediction. While measuring the prediction accuracy of recommendation to filter several recommendation approaches is important, it is not the only evaluation way of a certain recommendation approach. Nonperformance measures, such as serendipity, diversity, novelty, or coverage, can also evaluate recommendation approaches [13]. Only measures that used in this dissertation are discussed, namely: user coverage and diversity.

5.2.2.1 User coverage

One measure that compares different recommending algorithms in their capability to produce recommendations for larger set of users is the **user coverage** (U_{cov}). The more the coverage is, the better the recommending algorithm. U_{cov} is given by Equation (5.6):

$$U_{cov}=\frac{\sum_{u\in U} \rho_u}{|U|} \tag{5.6}$$

where $\rho_u = \begin{cases} 1 : if\ |recset_u| > 0 \\ 0 : \ else \end{cases}$

from the above equation, coverage measure is the ratio of users who receive nonempty recommended set to the total number of users [134]. Coverage is an important measure to analyze the recommender system in respect to new users with few known ratings.

5.2.2.2 Diversity Measure

Diversity is usually defined as the opposite of similarity [135]. In this dissertation, we calculated the diversity score of the list of recommended items to test whether the recommendations generated for users who are socially connected with indirect relations are more diverse than the recommendations that are produced using the direct social relations between users. One of the ways to measure the diversity of recommended items produced by different recommending algorithms is to measure the item-item similarity between one item and all of the other recommended items in the list. Then the sum, average, minimum, or maximum is calculated and compared to the corresponding value produced by another recommending algorithm. Smyth and McClave proposed one popular approach to calculate the diversity score [135], which is given by Equation (5.7):

$$\text{Diversity}\ (c_1 \ldots c_n) = \frac{\sum_{i=1..n} \sum_{\mathrm{j=i..n}} (1 - \text{Similarity}(c_i, c_j))}{\frac{n}{2} * (\mathrm{n} - 1)} \tag{5.7}$$

where $\text{Similarity}(c_i, c_j)$ is the similarity between item c_i and item c_j in the list of n recommended items.

5.3 Comparing Recommendation Approaches using Different Implicit Social Networks as Sources of Recommendations

Objective: The aim of this study is to compare the prediction accuracy of using different implicit SNs as sources of recommendation. Using different recommendation algorithms and different ISNs, we tested the prediction accuracy to explore which recommendation algorithm works the best for each ISN (experiment 1) and which ISN produces the best prediction accuracy (experiment 2). In addition, we tested the role of the neighbourhood size (K value) so that we could choose the best settings to be used in the next experiments. Figure 5.1 shows the design space of the first two experiments.

Data set used: For all of the experiments in this section, we used the proposed ISNs from Chapter 4. We built them from data that we crawled from CiteULike.

Dependent variables: Because the data we used are based on the users' bookmarking behaviours, we used the information retrieval evaluation measures with different ranks (number of recommendations given [i.e., top N]). We used precision, recall, and F1 measure at different ranks (N = 2, 5, 10).

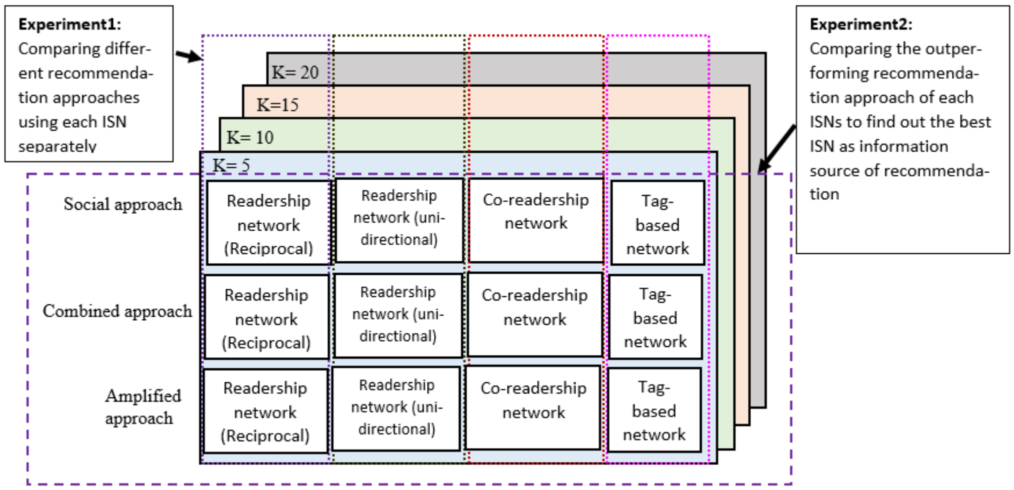


Figure 5-1: Design space of the first two experiments in this study

5.3.1 Experiment 1: Testing Different Recommendation Approaches for Each ISN

The main purpose of this experiment is to discover the best recommendation approach(es) for each implicit SN given a certain implicit SN and the set of recommendation approaches presented in section 5.1: social recommender, combined recommender, and amplified recommender. We also test different neighbourhood sizes to determine whether this factor affect the performance of the same recommender system using the same ISN. We select and use the best approaches and the size of the neighbourhood for each ISN in the next experiments. This experiment is shown in Figure 5.1 as columns. Each column compares different recommendation approaches for one of the ISNs. We need to answer the following research question:

RQ5.1: Using different social recommendation approaches, which approach works the best for each ISN?

To answer the previous research question, we tested eight hypotheses, two for each ISN, using the following independent variables: recommendation approaches (refer to section 5-1), different ISNs, and neighbourhood size (k = 5, 10, 15, 20). We discuss the hypotheses and the results for each ISN in the following sections.

We used descriptive statistics to summarize the data and explore normality. We used Kolmogorov-Smirnov (K-S) tests and histograms to check whether dependent variables (precision, recall, and F1 at ranks 2, 5, and 10) follow normal distribution. It appears that data are approximately normal (majority of p-values for the K-S test > 0.05), so using parametric statistical techniques (ANOVAs) is appropriate. We used the two-way factorial ANOVA to explore how algorithm (3 levels factor) and neighbourhood size (4 levels factor) affect prediction accuracy represented by precision, recall, and F1 measure.

5.3.1.1 Results for Reciprocal Readership ISN

To compare the prediction accuracy for different recommendation approaches using reciprocal readership ISN, we tested the following hypotheses:

> **$H0_{5.1}$**: There is no statistically significant difference between the means of *prediction accuracy measures* (i.e., precision, recall, F1 measure) at different ranks of different recommendation approaches using *reciprocal readership ISN* as the information source.
>
> **$H_{5.2}$**: The prediction accuracy of the three tested recommendation approaches based on *reciprocal readership ISN* outperform the conventional CF approach.

The results for testing hypothesis $H0_{5.1}$ are shown in Table 5.2; significant results are shown in boldface. The two-way ANOVA test shows that there is a statistically significant difference of the mean of prediction accuracy of recommendation approaches for all information retrieval measures (precision, recall, and F1) at all different ranks (N = 2, 5, 10); therefore, the null hypothesis $H0_{5.1}$ is rejected. However, there is no statistical difference of prediction accuracy for different neighbourhood size. We applied the Tukey post hoc test to check which recommendation approach has the highest prediction accuracy because the two-way ANOVA shows there is a statistical difference but it cannot show which one is the best. Tukey post hoc test is suitable to compare the results of different groups that have the same group sizes. The Tukey post hoc test shows that the social approach outperformed the combined and the amplified approaches but that there was no statistical difference between the prediction accuracy of the combined and amplified approaches; this was true for all prediction accuracy measures. Figure 5-2 depicts the comparisons of all three algorithms for all neighbourhood sizes for reciprocal readership ISN.

Table 5-2: Two-way ANOVA results for reciprocal readership ISN

Group	P@2	P@5	P@10	R@2	R@5	R@10	F1@2	F1 @5	F1@10
Social approach	.1706±.0172	.1424±.0115	.1023±.0073	.1056±.0143	.1979±.0150	.2588±.0149	.1303 ±.0152	.1654 ±.0112	.1472 ±.0089
Combined approach	.1457±.0108	.1227±.0010	.0904±.0045	.0924±.0119	.1739±.0182	.2299±.0202	.1127 ±.0108	.1435 ±.0113	.1295 ±.0060
Amplified approach	.1527±.0141	.1271±.0103	.0957±.0080	.1010±.0130	.1830±.0134	.2418±.0146	.1212 ±.0119	.1497±.0091	.1370 ±.0093
K = 5	.1585±.0185	.1325 ±.146	.0989±.0088	.0983±.0124	.1809±.0205	.2421±.0192	.1209 ±.0136	.1526 ±.0152	.1402 ±.0107
K = 10	.1597±.0159	.1319 ±.0111	.0956±.0075	.1022±.0158	.1890±.0206	.2445±.0238	.1244 ±.0158	.1551 ±.0128	.1373 ±.0105
K = 15	.1518±.0180	.1284 ±.0154	.0945±.0089	.0968±.0141	.1831±.0163	.2393±.0199	.1179 ±.0148	.1506 ±.0147	.1353 ±.0113
K = 20	.1554±.0184	.1299 ±.0134	.0965±.0090	.1014±.0142	.1867±.0165	.2479±.0194	.1223 ±.0145	.1530 ±.0139	.1388 ±.0116
Homogeneity of variance assumption (Levene's test)	p =.632	p =.494	p =.105	p =.452	p =.357	p = .441	p = .27	p = .661	p = .412
Main effect of approach	**F(2,54)= 16.114***	**F(2,54)= 18.357***	**F(2,54) = 17.405***	**F(2,54) =5.175***	**F(2,54) = 11.852***	**F(2,54) = 14.708***	**F(2,54) = 9.334***	**F(2,54) = 22.286***	**F(2,54) =23.607***
Main effect of neighborhood size (K)	F(3,54) = .918, p = .44	F(3,54) =.443 , p = .723	F(3,54) = 1.163, p = .332	F(3,54) =.563 , p = .642	F(3,54) =.798 , p = .50	F(2,54) = .695, p = .559	F(3,54) = .67, p = .57	F(3,54) = .445, p =.72	F(3,54) = .993, p = .40

* **Significant at p<0.01**

Next, we compared the three recommendation approaches to the conventional CF, which was used as a baseline. The *t*-test was used to compare each recommendation approach to the CF. Results of the *t*-tests are reported in Table 5-3; significant results are shown in boldface. The results show that the social approach and the amplified approach outperformed the CF significantly for all used measures except for P@5 of the amplified approach, where the CF and amplified approaches have the same performance. However, the combined approach performs similarly to CF without a significant difference. So, $H_{5.2}$ is partially accepted.

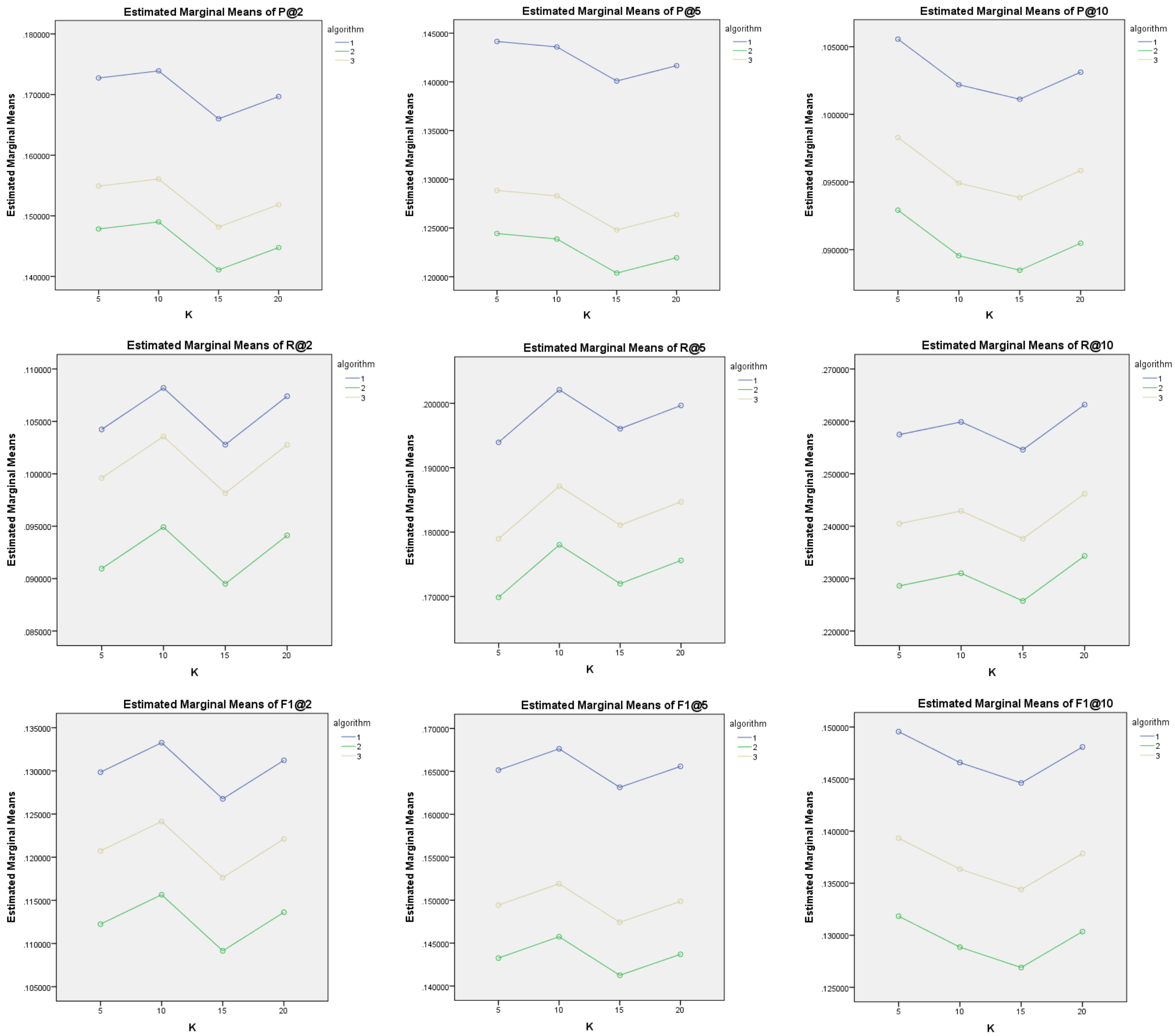


Figure 5-2: Comparisons of prediction accuracy of different recommendation approaches using data from reciprocal readership ISN (1-Social, 2- Combined, 3- Amplified)

Table 5-3: *T*-test results for comparing different recommending approach with collaborative filtering for reciprocal readership ISN

RS	Mean value (T-test value and significance level (p-value))		
	P@2	**P@5**	**P@10**
SR	**.1706 (t=5.775*)**	**.1424(t=6.040*)**	**.1029(t=5.775*)**
CR	.1457 (t=1.557, p=.128)	.1227 (t=.719, p=.477)	.0904(t=.393, p=.696)
AR	**.1527 (t=2.809*)**	.1270 (t=1.969, p=.056)	**.0957(t=2.532, p<.05)**
CF	.1383	.1202	.0896
	R@2	**R@5**	**R@10**
SR	**.1056 (t=4.017*)**	**.1979 (t=4.941*)**	**.2588 (t=5.671*)**
CR	.0924(t=.949, p=.349)	.1739(t=699, p=.489)	.2299(t=.258, p=.798)
AR	**.1010 (t=3.071*)**	**.1829 (t=2.422*)**	**.2417 (t=2.529*)**
CF	.0888	.1695	.2283
	F1@2	**F1@5**	**F1@10**
SR	**.1303 (t=4.778*)**	**.1654 (t=6.008*)**	**.1472 (t=6.251*)**
CR	.1127 (t=1.156, p=.255)	.1435 (t=.721, p=.475)	.1295 (t=.348, p.730)
AR	**.1212 (t=3.153*)**	**.1496 (t=2.361*)**	**.1370 (t=2.762*)**
CF	.1303	.1405	.1286

*** Significant at p<0.01**
SR: Social Recommender
CR: Combined Recommender
AR: Amplified Recommender
CF: Collaborative Recommender

5.3.1.2 Results for Unidirectional Readership ISN

We also tested the same hypotheses for reciprocal readership ISN for unidirectional readership as follows:

> **H0$_{5.3}$**: There is no statistically significant difference between the means of *prediction accuracy measures* (i.e., precision, recall, F1 measure) at different ranks of different recommendation approaches using the *unidirectional readership network* as the information source.

> **H$_{5.4}$**: The prediction accuracy of the three tested recommendation approaches based on the *unidirectional readership ISN* outperform the conventional CF approach.

The two-way ANOVA test results show that there is a statistically significant difference in prediction accuracy of difference recommendation approaches and also a statistically significant difference in prediction accuracy of different neighbourhood sizes (please refer to Table 5-4).

Therefore, the null hypothesis $H0_{5.3}$ is rejected. The Tukey post hoc test shows that the combined and the amplified approach outperformed the social approach, which is based on the pure social relations between users; results are true for all performance measures. However, the combined and the amplified approaches performed the same with no statistically significant difference between their prediction accuracy, and the combined approach has the higher value. The Tukey post hoc test, which was applied for different neighbourhood sizes, shows that the prediction accuracy when k = 10, 15, 20 is higher than the prediction accuracy when k = 5. However, there is no statistical difference between the prediction accuracy when k = 10, 15, or 20. The highest prediction accuracy is achieved when k = 20. Refer to Figure 5-3 for the comparisons of all three recommendation approaches across all neighbourhood sizes.

Table 5-4: Two-way ANOVA results for unidirectional readership ISN

Group	P@2	P@5	P@10	R@2	R@5	R@10	F1@2	F1 @5	F1@10
Social approach	.0054±.0007	.0047±.0005	.0041±.0004	.0019±.0002	.0038±.0006	.0062±.0007	.0027±.0022	.0042±.0005	.0049±.0004
Combined approach	.0657±.0044	.0509±.0023	.0407±.0017	.0227±.0014	.0408±.0019	.0615±.0027	.0338±.0021	.0453±.0020	.0489±.0019
Amplified approach	.0645±.0034	.0501±.0025	.0405±.0019	.0222±.0013	.0404±.0021	.0605±.0030	.0330±.0017	.0447±.0021	.0485±.0022
K = 5	.0416±.0264	.0329±.0205	.0268±.0165	.0148±.0095	.0268±.0168	.0405±.0251	.0218±.0139	.0296±.0185	.0323±.0199
K = 10	.0456±.0294	.0353±.0225	.0285±.0179	.0157±.0102	.0284±.0181	.0429±.0270	.0234±.0151	.0315±.0200	.0342±.0216
K = 15	.0468±.0305	.0359±.0230	.0290±.0183	.0160±.0104	.0289±.0185	.0438±.0275	.0238±.0155	.0320±.0205	.0349±.0219
K = 20	.0469±.0305	.0368±.0235	.0294±.0186	.0159±.0104	.0292±.0186	.0438±.0277	.0237±.0155	.0325±.0208	.0352±.0223
Homogeneity of variance assumption (Levene's test)	p =.018	p =.005	p =.001	p =.440	p =.349	p =.103	p =.476	p =.043	p =.043
Main effect of approach	**$F(2,54)$ = 4067.670***	**$F(2,54)$=6741.789***	**$F(2,54)$ =7322.430***	**$F(2,54)$ =2541.494***	**$F(2,54)$ =4436.080***	**$F(2,54)$ =5162.828***	**$F(2,54)$ =3330.990***	**$F(2,54)$ =6896.667***	**$F(2,54)$ = 7603.758***
Main effect of neighborhood size (K)	**$F(3,54)$ =16.189***	**$F(3,54)$ =18.769***	**$F(3,54)$ =15.460***	**$F(3,54)$ =4.185***	**$F(3,54)$ =8.233***	**$F(3,54)$ =9.741***	**$F(3,54)$ =7.151***	**$F(3,54)$ =15.361***	**$F(3,54)$ = 15.279***

* **Significant at p<0.01**

When we based the recommendation approaches on unidirectional readership ISN compared to the CF approach, the *t*-test shows that the precision values of the CF outperformed the precision values of the social approach (please refer to Table 5-5). The same pattern appears for the recall and F1 values. However, although the prediction accuracy of the combined and the amplified approaches is higher than the CF values, the difference is insignificant, which means that the combined and the amplified approaches are equivalent to the CF. Therefore, hypothesis $H_{5.4}$ is rejected.

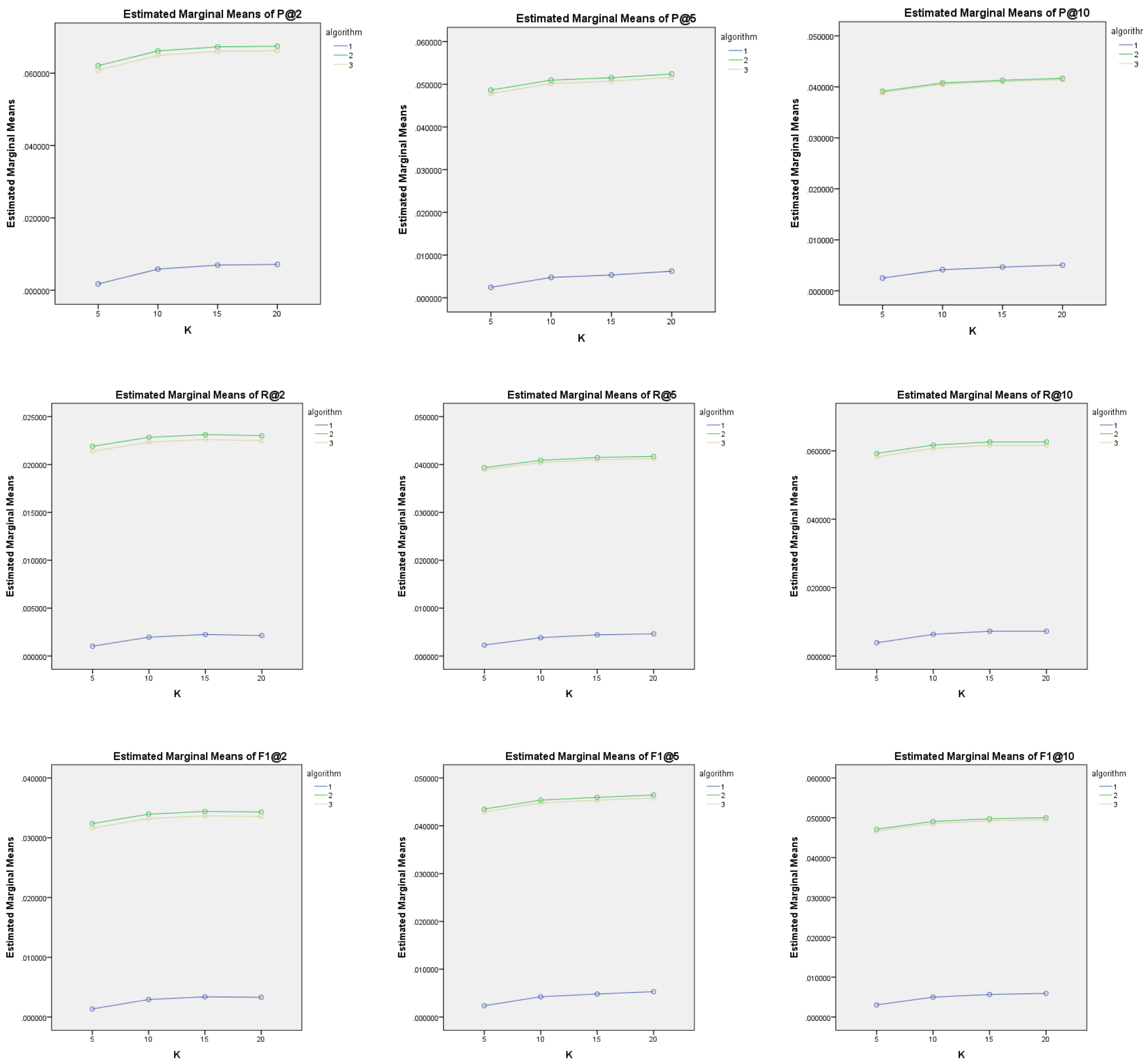


Figure 5-3: Comparisons of prediction accuracy of different recommendation approaches using data from unidirectional readership ISN (1-Social, 2- Combined, 3- Amplified)

Table 5-5: *T*-test results for comparing different recommending approach with collaborative filtering for unidirectional readership ISN

RS	Mean value (T-test value and significance level (p-value))		
	P@2	**P@5**	**P@10**
SR	**.0054(t=-75.805*)**	**.0047(t=-97.627*)**	**.0041(t=-100.406*)**
CR	.0657(t=1.125, p=.268)	.0509(t=.573, p=.570)	.0407(t=.151, p=.881)
AR	.0645(t=.151, p=.881)	.0501(t=-.550, p=.586)	.0405(t=-.295, p=.770)
CF	.0643	.0505	.0406
	R@2	**R@5**	**R@10**
SR	**0.0019(t=-69.122*)**	**.0038(t=-81.273*)**	**.0062(t=-93.744*)**
CR	.0227(t=1.742, p=.090)	.0408(=.502, p=.619)	.0615(t=.198, p=.844)
AR	.0222(t=.568, p=.573)	.0404(t=-.239, p=.812)	.0605(t=-.930, p=.358)
CF	.0219	.0405	.0613
	F1@2	**F1@5**	**F1@10**
SR	**.0027(t=-74.790*)**	**.0042(t=-94.156*)**	**.0049(t=-102.631*)**
CR	.0338(t=1.661, p=.105)	.0453(t=.568, p=.574)	0.489(t=.175, p=.862)
AR	.0330(t=.493, p=.625)	.0447(t=-.414, p=.681)	.0485(t=-.595, p=.556)
CF	.0327	.0449	.0489

*** Significant at p<0.01**
SR: Social Recommender
CR: Combined Recommender
AR: Amplified Recommender
CF: Collaborative Recommender

5.3.1.3 Results for Co-readership ISN

We tested the following hypotheses:

> **$H0_{5.5}$**: There is no statistically significant difference between the means of *prediction accuracy measures* (i.e., precision, recall, F1 measure) at different ranks of different recommendation approaches using the *co-readership ISN* as the information source.
>
> **$H_{5.6}$**: The prediction accuracy of the three tested recommendation approaches based on the *co-readership ISN* outperform the conventional CF approach.

As shown in Table 5-6, the two-way ANOVA test proves that there is a statistically significant difference of prediction accuracy of both recommendation approaches and neighbourhood size variables when co-readership ISN is used as a source of recommendation; this is true for all performance measures. Thus, the null hypothesis $H0_{5.5}$ is rejected. By applying the Tukey post hoc test, we found that the combined and the amplified approaches significantly outperformed the social approach. However, even though the combined approach has a higher prediction value than the amplified approach, the difference is insignificant, which means they have a similar performance for the co-readership ISN. The Tukey post hoc test for testing the differences between prediction accuracy for different neighbourhood sizes shows that the prediction accuracy when k = 10, 15, 20 is higher than the prediction accuracy when k = 5. However, there is no statistical difference between the prediction accuracy when k = 10, 15, or 20. For the full comparisons, refer to Figure 5-4.

Table 5-6: Two-way ANOVA results for co-readership ISN

Group	P@2	P@5	P@10	R@2	R@5	R@10	F1@2	F1 @5	F1@10
Social approach	.0415±.0031	.0316± .0019	.0251± .0013	.0166± .0008	.0287± .0014	.0423± .0017	.0237± .0013	.0301± .0015	.0315±.0015
Combined approach	.0592±.0021	.0451±.0016	.0353± .0012	.0242±.0010	.0425± .0019	.0628 ±.0028	.0343± .0014	.0438±.0017	.0452±.0017
Amplified approach	.0581±.0032	.0448±.0022	.0353±.0016	.0237±.0016	.0421± .0023	.0626± .0033	.0337± .0021	.0438±.0022	.0452±.0021
K = 5	.0489±.0087	.0378±.0065	.0299±.0049	.0199±.0034	.0284 ±.0049	.0365±.0064	.0380±.0065	.0284 ±.0049	.0365±.0064
K = 10	.0529±.0088	.0405±.0067	.0319±.0050	.0215±.0036	.0306±.0051	.0391±.0065	.0406±0066	.0306±.0051	.0391±.0065
K = 15	.0544±.0556	.0416±.0068	.0327±.0051	.0221±.0038	.0314±.0053	.0402±.0069	.0416±0069	.0314±.0053	.0402±.0069
K = 20	.0556±.0079	.0422±.0063	.0333±.0049	.0224±.0038	.0319±.0052	.0407±.0067	.0423±.0068	.0319±.0052	.0407±.0067
Homogeneity of variance assumption (Levene's test)	*p* =.877	*p* =.875	*p* =.229	*p* =.352	*p* = .338	*p* =-.772	*p* =.357	*p* =.338	*p* =-.772
Main effect of approach	***F*(2,54) =1521.949***	***F*(2,54) =2108.122***	***F*(2,54) =3282.969***	***F*(2,54) =667.036***	***F*(2,54) =977.945***	***F*(2,54) =2017.992***	***F*(2,54) =2982.982***	***F*(2,54) =977.945***	***F*(2,54) =2017.992***
Main effect of neighborhood size (K)	***F*(3,54) =98.506***	***F*(3,54) =101.560***	***F*(3,54) =158.128***	***F*(3,54) =31.201***	***F*(3,54) = 50.092***	***F*(3,54) =87.381***	***F*(3,54) =128.469***	***F*(3,54) = 50.092***	***F*(3,54) =87.381***

* **Significant at p<0.01**

We also compared the results of the recommendation approaches (social, combined, and amplified) to the baseline approach (CF). The *t*-test shows that the CF approach outperformed the social approach significantly using all performance measures (please refer to Table 5-7). The combined approach outperformed the CF significantly for P@2. Although for all performance measures the combined and the amplified approaches have higher prediction values than the CF, the difference is insignificant. Therefore, $H_{5.6}$ is partially accepted.

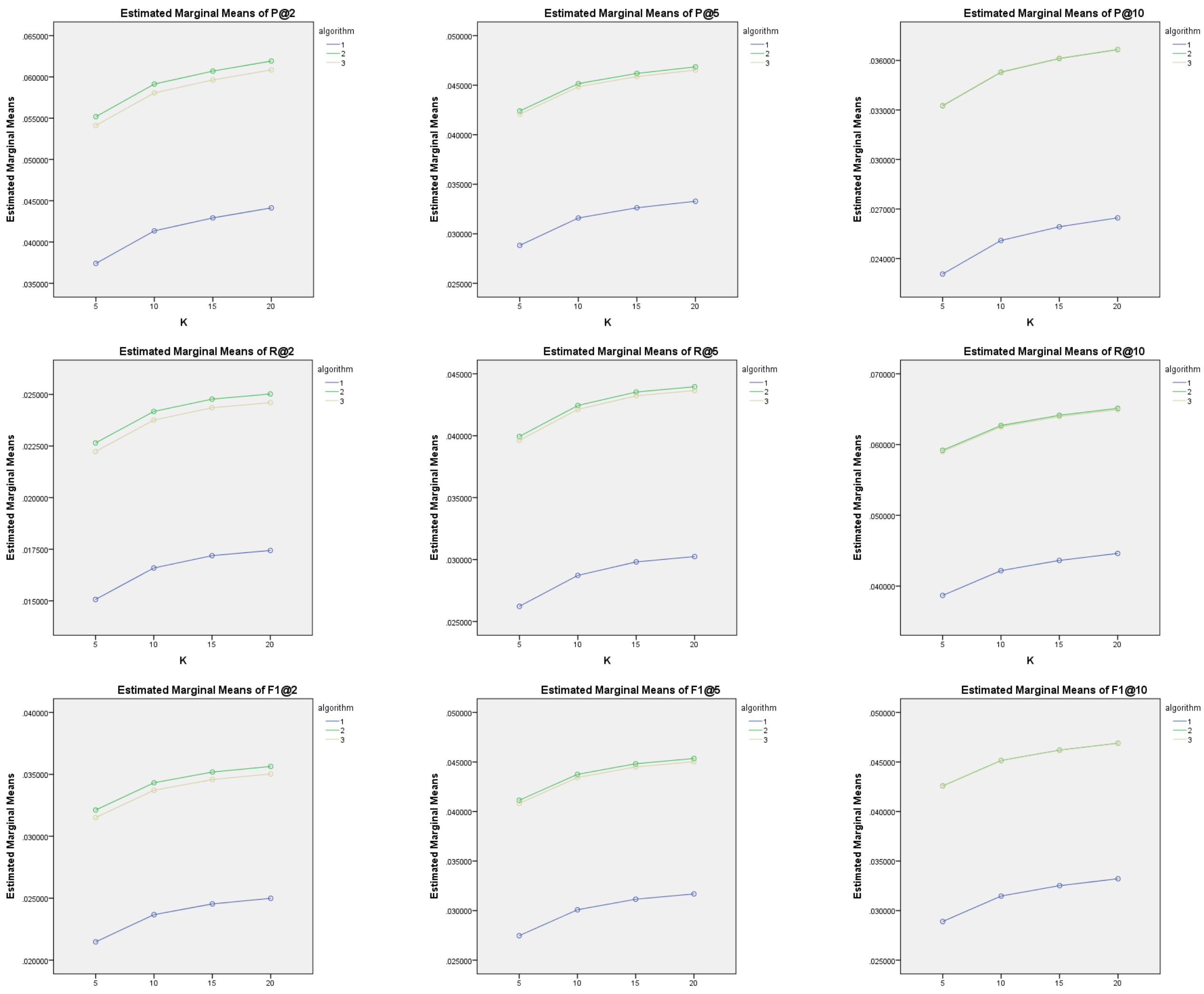


Figure 5-4: Comparisons of prediction accuracy of different recommendation approaches using data from co-readership ISN (1-Social, 2- Combined, 3- Amplified)

Table 5-7: *T*-test results for comparing different recommending approach with collaborative filtering for co-readership ISN

RS	Mean value (T-test value and significance level (p-value))		
	P@2	**P@5**	**P@10**
SR	**.0415(t=-16*)**	**.0316(t=-20.433*)**	**.0251(t=-22.216*)**
CR	**.0592(t=2.223, p<.05)**	.0451(t=1.637, p=.11)	.0353(t=1.518, p=.137)
AR	.0581(t=.834, p=.410)	.0448(t=.942, p=.352)	.0354(t=1.392, p=.172)
CF	.0573	.0442	.0347
	R@2	**R@5**	**R@10**
SR	**.0165(t=-17.231*)**	**.0287 (t=-20.504*)**	**.0423(t=-23.194*)**
CR	**.0241(t=2.185, p<.05)**	.0425(t=1.698, p=.098)	.0628(t=1.879, p=.068)
AR	.0237(t=.977, p=.335)	.0422(t=1.154, p=.256)	.0626(t=1.545, p=.131)
CF	.0233	.0413	.0610
	F1@2	**F1@5**	**F1@10**
SR	**.0236(t=-17.497*)**	**.0301(t=-21.103*)**	**.0315(t=-23.248*)**
CR	.0343(t=2.238, p<.05)	.0438(t=1.703, p=.097)	.0452(t=1.702, p=.097)
AR	.0337(t=.956, p=.345)	.0434(t=1.082, p=.286)	.0452(t=1.483, p=.146)
CF	.0331	.0427	.0442

*** Significant at p<0.01**
SR: Social Recommender
CR: Combined Recommender
AR: Amplified Recommender
CF: Collaborative Recommender

5.3.1.4 Results for Tag-based ISN

We tested the following hypotheses:

$H0_{5.7}$: There is no statistically significant difference between the means of *prediction accuracy measures* (i.e., precision, recall, and F1 measure) at different ranks of different recommendation approaches using the *tag-based ISN* as the information source.

$H_{5.8}$: The prediction accuracy of the three tested recommendation approaches based on the *tag-based ISN* outperform the conventional CF approach.

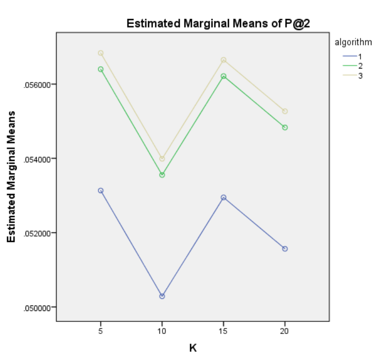
The two-way ANOVA test results (please refer to Table 5-8) show that there is a statistically significant difference between the performance of different recommendation approaches for most of the measures except P@2, F1@5, and F1@10, where the differences are insignificant but the amplified approach had the highest prediction accuracy. Therefore, the null hypothesis $H0_{5.7}$ is rejected. The Tukey post hoc test shows that for P@5 and P@10, the social recommender outperformed the combined and amplified approaches significantly; however, there were no significant differences between the combined and the amplified approaches' performances. Across other measures, the amplified and the combined approaches outperformed the social approach significantly, but there was no statistical significance between the combined and the amplified approaches' performances. The highest prediction value was achieved by the amplified approach. In addition, the Tukey post hoc test shows that there was no statistically significant difference in the prediction accuracy of different neighbourhood sizes. Figure 5-5 shows the comparisons of different algorithms across different neighbourhood sizes when tag-based ISN was used.

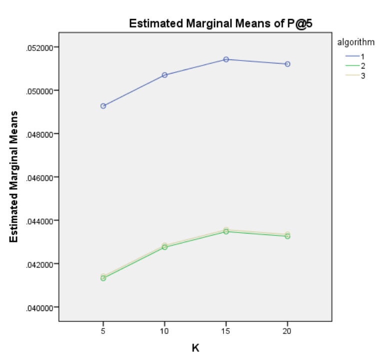
Table 5-8: Two-way ANOVA results for tag-based ISN

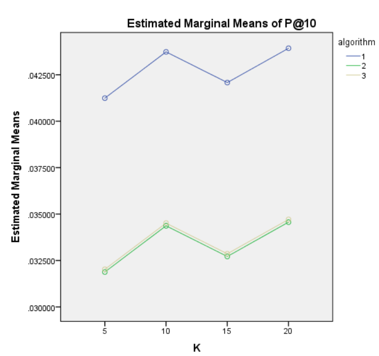
Group	P@2	P@5	P@10	R@2	R@5	R@10	F1 @ 2	F1 @ 5	F1 @ 10
Social approach	.0520±.0160	.0507±.0090	.0427±.0094	.0122±.0077	.0313±.0176	.0487±.0188	.0192±.0107	.0366±.0137	.0435±.0097
Combined approach	.0552±.0033	.0427±.0021	.0334±.0015	.0214±.0016	.0385±.0022	.0573±.0032	.0308±.0022	.0405±.0021	.0422±.0021
Amplified approach	.0557±.0032	.0428±.0019	.0335±.0015	.0217±.0016	.0388±.0022	.0575±.0030	.0311±.0021	.0407±.0021	.0424±.0020
K = 5	.0555±.0127	.0440±.0078	.0350±.0068	.0183±.0061	.0345±.0112	.0502±.0120	.0271±.0083	.0376±.0095	.0402±.0068
K = 10	.0526±.0099	.0454±.0068	.0375±.0083	.0176±.0064	.0339±.0087	.0543±.0101	.0260±.0085	.0378±.0068	.0433±.0051
K = 15	.0553±.0074	.0462±.0066	.0359±.0061	.0188±.0061	.0359±.0094	.0546±.0112	.0277±0079	.0394±.0071	.0420±.0030
K = 20	.0539±.0083	.0459±.0051	.0377±.0070	.0188±.0074	0405±.0129	.0590±.0128	.0275±.0097	.0422±.0089	.0452±.0067
Homogeneity of variance assumption (Levene's test)	p =.005	p =.005	p =.005	p =.005	p =.005	p =.005	p =.005	p =.005	p =.005
Main effect of approach	$F(2,54)$ =.850, p =.433	**$F(2,54)$ =13.708***	**$F(2,54)$ =18.372***	**$F(2,54)$ =25.631***	**$F(2,54)$ =3.447***	**$F(2,54)$ =4.177***	**$F(2,54)$ =21.261***	$F(2,54)$ =1.613, p =.209	$F(2,54)$ =.319, p =.728
Main effect of neighborhood size (K)	$F(3,54)$ =.276, p = .843	$F(3,54)$ =.462, p =.710	$F(3,54)$ =.809, p =.494	$F(3,54)$ =.205, p =.892	$F(3,54)$ =1.305, p = .282	$F(3,54)$ =1.589, p = .203	$F(3,54)$ =.197, p = .898	$F(3,54)$ =1.065, p = .372	$F(3,54)$ =1.996, p = .125

* **Significant at p<0.01**

Furthermore, we compared the performances of the three recommendation approaches (social, combined, and amplified) to the CF approach. Table 5-9 represents the results of the *t*-test, which show that the social approach's precision at 5 and at 10 outperformed the CF precision values for the same ranks but that the CF approach recall and F1 values were significantly higher than the social approaches' when the rank was low (i.e., R@2, F1@2). For the other measures, the amplified approach had higher values than the CF approach, but the difference is insignificant. Thus, hypothesis $H_{5.8}$ is partially accepted.







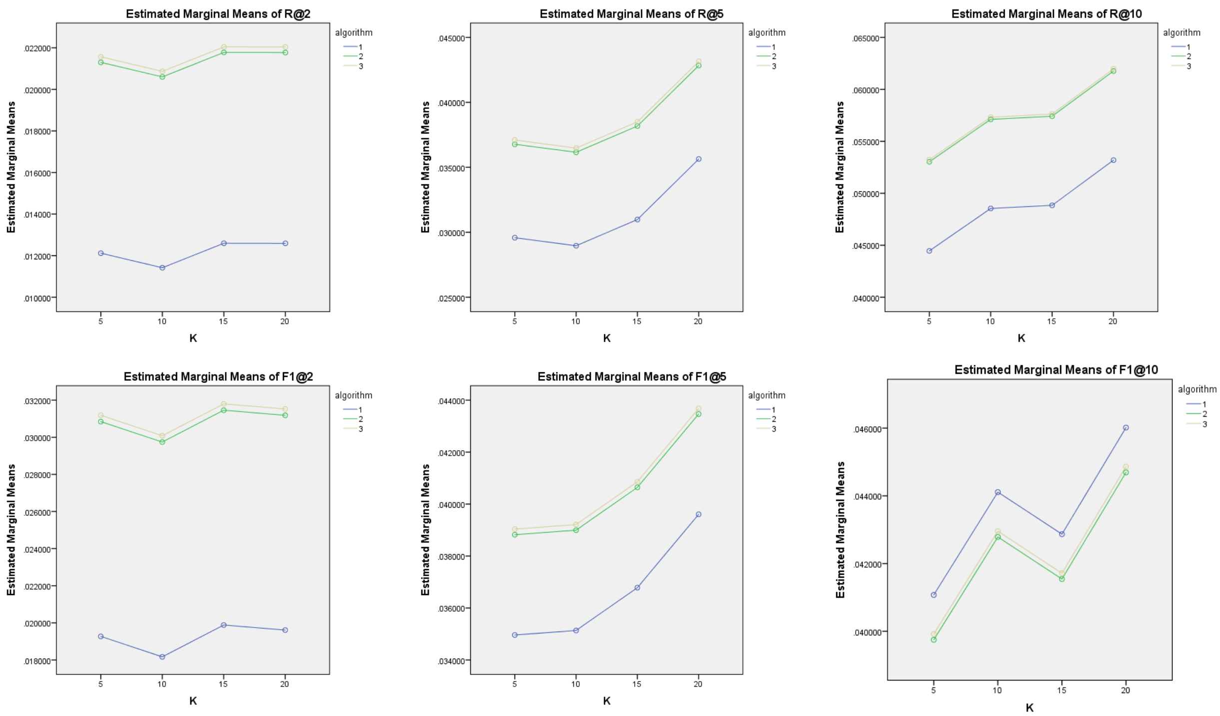


Figure 5-5: Comparisons of prediction accuracy of different recommendation approaches using data from tag-based ISN (1-Social, 2- Combined, 3- Amplified)

Table 5-9: *T*-test results for comparing different recommending approach with collaborative filtering for tag-based ISN

RS	Mean value (T-test value and significance level (p-value))		
	P@2	**P@5**	**P@10**
SR	.0520(t=-.896, p=.376)	**.0507(t=3.893*)**	**.0427(t=4.374*)**
CR	.0552(t=-.014, p=.989)	.0427(t=.139, p=.890)	.0334(t=-.062, p=.951)
AR	.0557(t=.402, p=.690)	.0428(t=.275, p=.785)	.0335(t=.232, p=818)
CF	.0552	.0426	.0334
	R@2	**R@5**	**R@10**
SR	**.0128(t=-5.163*)**	.0313(t=-1.837, p=.074)	.0488(t=-2.071, p=.045)
CR	.0214(t=.118, p=.906)	.0385(t=-.142, p=.888)	.0573(t=-.257, p=.798)
AR	.0216(t=.628, p=.534)	.0389(t=.295, p=.769)	.0575(t=-.063, p=.950)
CF	.0213	.0385	.0576
	F1@2	**F1@5**	**F1@10**
SR	**.0192(t=-4.7*)**	.0366(t=-1.247, p=.220)	.0435(t=.551, p=.585)
CR	.0308(t=.091, p=.928)	.0405(t=-.022, p=.983)	.0423(t=-.143, p=.887)
AR	.0311(t=.582, p=.564)	.0407(t=.289, p=.774)	.0424(t=.112, p=.911)
CF	.0307	.0404	.0422

*** Significant at p<0.01**
SR: Social Recommender
CR: Combined Recommender
AR: Amplified Recommender
CF: Collaborative Recommender

In the previous sections (5.3.1.1 through 5.3.1.4), we tested the recommendation approaches for each ISN with different neighbourhood sizes, aiming to answer the following research question:

Which recommendation approach works the best for each implicit social network?

To decide which setting works the best for each ISN, we considered the significant results of the two-way ANOVA test and the *t*-test. When the results were insignificant, we chose the dominant approach which has the highest value considering different measures.

Thus the following settings for each ISN is used for the next experiments:

- Reciprocal readership ISN: social approach with K=10
- Unidirectional readership ISN: combined approach with K=20
- Co-readership ISN: combined approach with K=20
- Tag-based ISN: amplified approach with K=20

5.3.2 Experiment 2: Testing Which ISN Delivers the Best Prediction Accuracy

The aim of this experiment is to explore which source of information (i.e., implicit SN) delivers the best recommendation quality given a certain implicit SN and given the outperforming recommendation approach from experiment 1 (section 5.3.1). This experiment is shown in Figure 5-1 as the big dashed purple box and it answers the following research question:

> **RQ5.2:** Comparing the recommendations using different ISNs, which one produces the highest prediction accuracy?

To answer this question, we tested the following null hypothesis:

> **$H0_{5.9}$:** There is no statistically significant difference between the means of prediction accuracy measures (i.e., precision, recall, F1 measure) at different ranks using different ISNs as information sources of recommendation.

The one-way ANOVA test was used to test the null hypothesis $H0_{5.9}$, which shows that there is a statistically significant difference between the means of all performance measures of the recommendations that are based on different ISNs: $F(3,16) = 305.540$ for P@2, $F(3,16) = 731.072$ for P@5, $F(3,16) = 429.922$ for P@10, $F(3,16) = 264.141$ for R@2, $F(3,16) = 420.272$ for R@5, $F(3,16) = 757.708$ for R@10, $F(3,16) = 363.589$ for F1@2, $F(3,16) = 880.790$ for F1@5, $F(3,16) = 674.972$ for F1@10. All results are significant at $p < 0.0005$. So, the null hypothesis $H0_{5.9}$ is rejected.

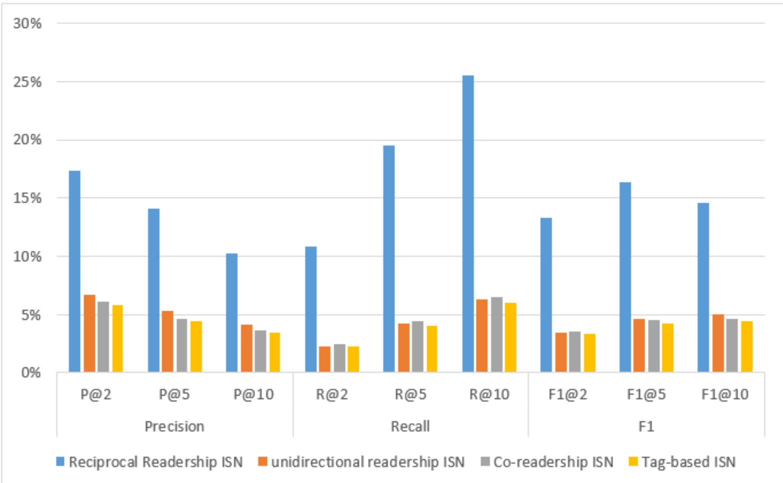


Figure 5-6: Comparison between of the performance of recommendations based on different ISNs

The Tukey post hoc test was then applied for the multiple comparisons because the one-way ANOVA test did not show which of the specific groups differed. The results of the Tukey post hoc test revealed that the highest value was achieved by the recommendations that use information from the reciprocal readership ISN as a source of recommendation, followed by recommendations based on unidirectional readership ISN, then the recommendation using co-readership ISN, and concluding with the recommendation based on tag-based ISN. However, there was no statistically significant difference between the recommendations based on unidirectional readership ISN or co-readership ISN, please refer to Figure 5-6.

5.4 Hybrid Recommendation using Recommendation from ISNs and Explicit SNs

Objective: The aim of this experiment was to test whether fusing different recommendations using data from ISNs and data from explicit SNs performs better than the performance of each recommendation alone. The outperformed recommendation approach for each ISN from experiment 1 was used in this experiment (experiment 3). For each explicit SN, friendship and co-authorship, we used the same approach to test which algorithm performed best exactly as we did for the ISNs.

We found that the best performing settings for each network with respect to more measures are:

- Readership ISN (reciprocal relations): social recommender with K=10
- Readership ISN (unidirectional relations): combined recommender K= 20
- Co-readership ISN: combined recommender, K= 20
- Tag-based ISN: amplified recommender, K= 20
- Co-authorship SN: amplified recommender, K= all relations
- Friendship SN: amplified recommender, K= all relations

Therefore, we used these settings in the next experiment when fusing data from different social networks. With compatible results with the study in [4], small neighborhood size provided the best accuracy results. In addition, as noted for the explicit social networks (friendship and co-authorship), the best results were achieved by using all the of users' social relations. This is because each user has very few social relations; the average number of relations in friendship networks and co-authorship networks are 0.3 and 1.27 respectively.

The question that we aimed to answer is

RQ5.3: Does fusing recommendations from ISNs and explicit SNs improve the performance of the recommendation?

First, we discuss the way that the recommendation from ISNs are fused with the recommendation of explicit SNs. Then the results of hybrid recommendation are discussed.

5.4.1 Finding the Best Weight Combination to Combine Recommendation from ISNs and explicit SNs

First, we used a weighted hybrid recommender to combine the results of recommending research papers using data from explicit and implicit social networks. Even though there are many hybrid approaches [16], we prefer to use the weighted hybrid approach because it brings together all the capabilities of the combined approaches in a straightforward and easy to perform way. It is a linear combination that aggregates the prediction score from different recommendation approaches using a different weight for each recommendation approach. The hybrid recommendation is calculated from the linear combination of different recommendations using the following equation:

$$Wrec_i = \sum_{S_j \in S} (W_{rec_{i,S_J}} \, . W_{S_J}) \tag{5.8}$$

where W_{S_J} is the weight for the recommender $S_{j,}$ and its value ranging from 0.1 to 0.9, and the sum of all weights is equal to 1. The optimum weight is usually derived by examining the performance of all possible combinations [16], then for the next recommendation, the weights are chosen so that the previous performance is retained. For example, the prediction accuracy of the new weight combination should be better or at least the same as the previous prediction accuracy. We used all the combinations from 0.1 to 0.9 by gradually increasing the weight of the first recommender by increments of 0.1. We first tested the hybrid approach of the co-authorship network (explicit) with every implicit social network, then we tested the friendship network (explicit) with every implicit social network. However, we used a modified version of the weighted sum approach called cross-source hybrid [121]. Cross-source hybrid approach favors items that are recommended by both approaches. Items that are recommended by implicit social network recommender and explicit social network recommender are more important than items that are recommended by only one recommender. Therefore, the above equation for weighted sum hybrid approach is modified as follows:

$$Wrec_i = \sum_{S_j \in S} (W_{rec_{i,S_J}} \, . W_{S_J}) \, . \left|S_{rec_i}\right| \quad (5.9)$$

Where $\left|S_{rec_i}\right|$ is the number of recommenders that recommend item i. We use the cross-source hybrid if the user has relations in both social networks. However, we used weight 1 for the recommendation if the user has relations in only one of social networks. For instance, if we aim to fuse the recommendations produced by co-authorship explicit network with recommendation produced by co-readership ISN, but the user has only relations in co-readership, we use the weight 1 for the recommendation produced by co-readership and completely ignore the co-authorship ISN for this specific user. We used this approach to make the recommendations more personalized. The best weight combination for each hybrid approach is shown in Table 5.10. When recommendations using co-authorship network are fused with recommendations from reciprocal readership ISN, the maximum accuracy is reached when the recommendations from the co-authorship network are given high weight, 0.8. However, when co-authorship network recommendation is fused with other ISNs, the best accuracy achieved when the weight of co-authorship was 0.3 in the case of unidirectional readership ISN and 0.1 in the case of co-readership ISN and tag-based ISN. This is because there is a high overlap between the co-authorship social relations and the reciprocal readership relations; 58.68 percent of the relations in the co-authorship network was discovered by the reciprocal relations in readership network.

The effect of the co-readership is less visible in the other networks, and that might be because there is a huge gap between the small number of relations in co-authorship network and the large number of relations in the other networks.

When recommendations from friendship network are fused with recommendations from implicit social networks, we can notice that the maximum accuracy of the recommendations occurs when the weight of the friendship network is higher than the weight of implicit networks.

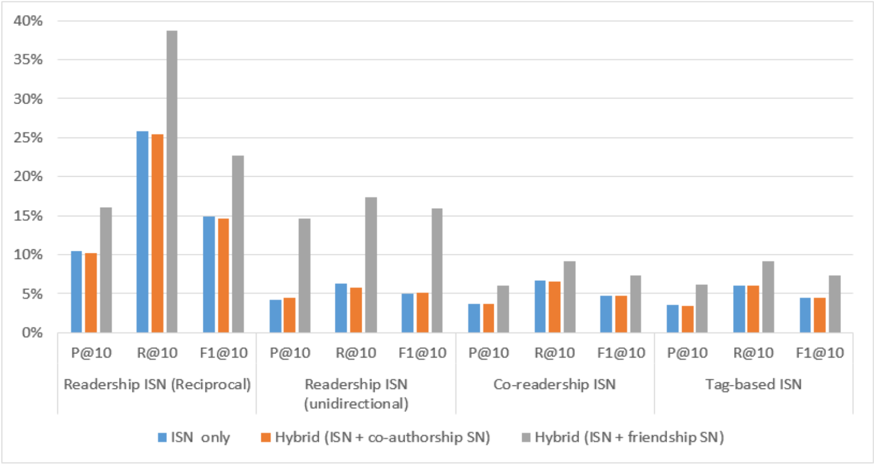
Table 5-10: The optimum weights for each hybrid approach (ISN weight, explicit SN weight)

	Co-authorship	**Friendship**
Readership ISN (Reciprocal)	(0.2,0.8)	(0.3,0.7)
Readership ISN (unidirectional)	(0.7,0.3)	(0.3,0.7)
Co-readership ISN	(0.9,0.1)	(0.1,0.9)
Tag-based ISN	(0.9,0.1)	(0.1,0.9)

The results of the best weight combinations are used in the next experiment described in the next section.

5.4.2 Comparison between Recommendations from Different ISNs with Friendship or Co-authorship SNs

We conducted an experiment to compare the recommendation using only ISN data, with the hybrid approach that combine recommendations from ISNs with each of the two explicit networks – the co-authorship network, or the friendship network, respectively. The results shown in Figure 5-7 reveal that the best prediction accuracy is achieved when the recommendation from the friendship network is fused with recommendation from ISN; this is true for all implicit social networks and for all measures (precision, recall and F1 at top ten). However, the co-authorship ISN did not help in increasing the prediction accuracy. In most of the cases, the prediction



accuracy stayed the same. In the readership ISN, the prediction accuracy slightly decreased when recommendation from co-authorship SN is fused. The only case that the precision increased is when recommendation from unidirectional readership SN is used with co-authorship SN recommendation.

Figure 5-7: Comparison between using recommendations from ISNs only, or fusing the recommendation with co-authorship SN and friendship network

5.4.3 User Coverage

We found that the co-readership ISN had the highest user coverage (87.25 percent), then the tag-based ISN (85.55 percent), the unidirectional readership ISN (37.22 percent), and the last is the reciprocal readership ISN (1.59 percent). We found that only 18 percent of users have explicit social relations and the average number of social relations per user is only 0.31. The co-authorship explicit SN has a very low coverage (1.873 percent). A tradeoff is noticed between the prediction accuracy and the user coverage: the more accurate the prediction, the smaller the user coverage.

Table 5-11 shows the coverage of different social networks and compares them to the coverage of the hybrid approaches. The recommendation coverage increases when recommendations from explicit and implicit SNs are combined. However, the maximum coverage is reached when recommendation from the friendship SN is fused with any of the ISNs. This is true for all of ISNs. For example, the increase in the coverage for the reciprocal readership ISN when the recommendation is fused with the friendship SN is more than 16 percent, while the increase in the coverage when the recommendation fused with the co-authorship SN is only 0.59 percent.

Fusing recommendation from friendship SN increases both the prediction accuracy and the recommendation coverage. The unidirectional readership ISN is the network that improved the most from fusing recommendation from friendship SN (F1-measure increase of almost 11 percent), then the reciprocal readership ISN (7.8 percent), tag-based ISN (2.9 percent increase), and finally, the co-readership ISN (with 2.5 percent increase). Even though, fusing recommendation from co-authorship SN did not improve the recommendation accuracy, it improves the recommendation coverage. It is important also to note that the user coverage for both co-authorship SN and friendship SN increased when the recommendation from each of them fused with each ISNs. The user coverage for co-authorship and friendship SNs are 1.87% and 18 % respectively.

Table 5-11: Comparison between the user coverage of using different hybrid approaches and using recommendation from ISNs alone

ISN	ISN data only	Hybrid with co-authorship SN	Hybrid with friendship SN
Reciprocal readership ISN	1.59%	2.18%	18.58%
Unidirectional readership ISN	37.22%	37.43%	45.14%
Co-readership ISN	87.25%	88.9%	92.71%
Tag-based ISN	85.55%	85.62%	86.57%

5.5 Summary

In this chapter, we measured the prediction accuracy and user coverage of the recommendation based on different ISNs. We used three different recommendation algorithms and four different neighbourhood sizes. First, we extensively studied the role of different algorithms and neighbourhood sizes on prediction accuracy using methods from the information retrieval field. We found that in most of the cases there was no statistical difference in the performance using different neighbourhood sizes. We also found that the hybrid approaches that use data from CF neighbours and social friends outperformed the pure social approach significantly in the unidirectional readership ISN, co-readership ISN, and tag-based ISN. Then we checked which ISN produced the best prediction accuracy, and we also measured the user coverage. The reciprocal readership ISN had the highest prediction accuracy and the lowest user coverage. Next, the unidirectional readership ISN and the co-readership ISN, which have similar prediction accuracy but higher user coverage for the co-readership ISN. Finally, the tag-based ISN has the lowest prediction accuracy with high user coverage.

In the last experiment, we compared fusing recommendations from ISNs with recommendations from two explicit SNs: friendship and co-authorship. The experiments show that fusing the recommendations from each ISN with recommendations from either the friendship or co-authorship explicit network is beneficial in increasing the user coverage. In addition, the prediction accuracy of all the recommendations from ISNs improved when fused with the friendship explicit SN, but fusing with the co-authorship SN did not help in improving the recommendation accuracy.

In this chapter, we studied the relationship between the prediction accuracy and user coverage because we aimed to find the approach that gives the best balance between the two measures. In Chapter 6, we study the balance between the prediction accuracy and the recommendation diversity. The majority of the previous studies in recommending research papers have focused on increasing the prediction accuracy by developing new algorithms. By doing so, the recommended list usually has items that are more relevant to the target user's interest, which means the items are similar to each other. However, relevant and diverse items are also important to the user. The diversity of the recommended items is an important but less studied nonperformance measure. Chapter 6 is dedicated to studying the relationship between the prediction accuracy and the diversity of the recommended list that is produced using data from the ISN.

CHAPTER 6:
IMPACT OF DIFFERENT SOCIAL DISTANCE LEVELS ON RECOMMENDATION

The recommendation performances for the different ISNs were discussed in the previous chapter; specifically, we compared only recommendations from data from ISNs that used a hybrid approach (combining recommendations from ISNs and two different explicit SNs) based on performance (i.e., prediction and accuracy at different ranks) and nonperformance (i.e., user coverage) measures. In this chapter, we explore the recommendations involving direct and indirect user relationships in each of the proposed ISNs; we want to explore the effects of indirect relationships on recommendations because direct relationships typically connect users more strongly than indirect relationships. Thus, we hypothesize that recommendations from distant friends can provide more diverse recommendations than recommendations from direct friends.

Most studies on recommender systems focus on increasing prediction accuracy by proposing different algorithms, where accuracy refers to identifying items that are more suited to user preferences or items that have interested users in the past. Providing users with more accurate lists, however, is not necessarily satisfactory; greater prediction accuracy means more items that are similar to users' preferences. With research papers, for example, a list of recommended papers by a single author whose work the user had read in the past is not necessarily a good list [136]; lists that recommend similar items may not be useful because users need more time to explore these similar items [13] and may miss other useful papers by recommending papers that are only similar in one aspect. Ziegler et al. suggested that recommendation lists be judged for diversity in their entirety rather than treating each recommended item as an isolated entity [137]; however, diversity in recommendation lists remains highly unexplored. In this chapter, we did not implement a new algorithm that increased diversity; rather, we measured the diversity of recommendation lists that considered direct and indirect relationships using different ISNs and then compared the results to identify the dominant curve after we plotted the precision-diversity curves of the recommendations using different social relation distances as suggested by [13]. This chapter answers the following research question:

RQ5.4: What is the effect of social relation distance on both prediction accuracy and diversity in recommendation lists?

First, we tested the same recommendation approaches that we used in chapter 5—social, combined and amplified—to determine which approach worked best for each network using indirect user relationships, and we used the same methodology to test the best settings using different neighbourhood sizes. The results for each ISN are discussed in the next section.

6.1 Results of Comparing Different Recommendation Approaches

6.1.1 Results for Reciprocal Readership ISN:

We conducted two-way ANOVA to test the following null hypotheses:

H0$_{6.1}$: There is no statistically significant difference between the means of *prediction accuracy measures* (i.e., precision, recall, F1 measure) at different ranks of recommendation approaches using the one-hop social relation distance and the *reciprocal readership ISN* as the information source.

H$_{6.2}$: The prediction accuracy of the three tested recommendation approaches based on the one-hop distance in the *reciprocal readership ISN* outperform the conventional CF approach.

For the one-hop social distance, the results for the differences between the means of all measures were significant for different recommendation approaches but not for different neighborhood sizes, and thus, the null hypothesis H0$_{6.1}$ was rejected. We applied the Tukey post hoc test to check which approach outperformed the others, and the results show that the social approach outperformed the combined and amplified approaches, although there was no statistically significant difference between the combined and amplified approaches themselves. To test hypothesis H$_{6.2}$, we used *t*-tests and found that the social approach significantly outperformed the CF approach but that the CF, combined and amplified approaches all had equivalent performance.

We tested the same hypotheses for the two-hop social distances, and two-way ANOVA showed that the only significant results were achieved when the number of recommended items was 10; there were statistically significant results in the mean prediction accuracy values for the different recommendation approaches for P@10, R@10 and F1@10. Tukey's post hoc test showed that the amplified and combined approaches significantly outperformed the social approach and that the amplified approach had a higher mean than the combined approach, although this difference was not significant.

The *t*-test we used to compare the results with the conventional CF approach showed that the CF significantly outperformed the social approach. However, the combined and amplified approaches performed the same as the CF in that the results were not significant.

6.1.2 Results for Unidirectional Readership ISN

The results for both indirect social relation distances (one and two hops) in the unidirectional readership ISN were compatible with each other. The performances of the amplified and combined approaches were significantly higher than that of the social approach, and that of the amplified approach was higher than that of the combined approaches; however, the differences in the means for all measures were not significant. In addition, the two-way ANOVAs that compared the results based on neighborhood size showed statistically significant differences in the means. The Tukey post hoc test showed that the means when K = 20 or 15 were significantly higher than the means when K = 10 or 5. In addition, the means when K = 10 was significantly higher than that when K = 5, and there was no statistically significant difference between the means when K = 20 versus K = 15.

The *t*-tests also showed that the results for the one-hop social distance were compatible with the results for the two-hop distance: The CF significantly outperformed the social approach. The amplified and combined approaches did have higher means than the CF, but the differences were not significant.

6.1.3 Results for Co-readership ISN

We also found that the results for both indirect social relation distances (one hop and two hops) for the co-readership ISN were compatible; the combined approach had a higher mean than the amplified approach, but the difference was not significant. However, both the combined and amplified approaches performed significantly better than the social approach. The two-way ANOVAs showed statistically significant differences in mean prediction accuracy by neighborhood size, and a follow-up Tukey test showed that when K = 20 or K = 15, the difference was significantly greater than when K = 10 or K = 5.

We used the *t*-test to compare the CF with each of the other algorithms, and the results for social distance of one hop were compatible with those for two hops; specifically, the CF significantly outperformed the social approach, and the amplified and combined approaches had higher means than the CF, but the differences were not significant.

6.1.4 Results for Tag-based ISN

The results for the tag-based ISN were completely compatible between one hop and two hops. The two-way ANOVA results showed statistically significant differences in mean prediction accuracy by recommendation approach and also by neighborhood size. We used the Tukey post hoc test to check which approach and which neighborhood size gave the best results and found that the combined and amplified approaches significantly outperformed the social approach but that there was no statistically significant difference in mean prediction accuracy between the combined and amplified approaches. In addition, mean prediction accuracy was higher when K = 20, 15 or 5 than when K = 10. There was no statistically significant difference when K = 20, 15 or 5, but the highest mean was achieved when K = 20.

The *t*-test showed that the CF significantly outperformed the social approach in precision and recall but that performance was the same when the F1 measure was considered, and these findings were true for both the one-hop and two-hop social distances. The *t*-test also showed that the combined, amplified and CV approaches all had equal performance at both social distances.

6.1.5 Choosing the Best Settings for Each ISN

Based on the results from the previous section, we can summarize that the social approach, which uses data from social peers, performed better than the CF approach and the two hybrid approaches (combined and amplified) in only one case: with the one-hop distance in the reciprocal readership ISN; it appears that the social approach worked best when the similarity between users was high and also when users were connected directly through the relations in the reciprocal readership ISN. For the other ISNs, the combined and amplified approaches outperformed the social approach but were similar to the CF approach results for either direct or indirect relationships.

To select the best settings for each ISN at different social distances, we chose the approach and neighborhood size that performed significantly better than the others considering the two-way ANOVA and *t*-test results. If there was no significant difference, we chose the approach with the highest values in most cases (i.e., using different prediction accuracy measures with different ranks). Table 6-1 shows the selected settings for each ISN case, and these selected approaches were used for the next experiment.

Table 6-1: The best approach for each ISN considering the social distance

ISN	**Setting for One hop social distance**	**Settings for Two hops social distance**
Reciprocal readership ISN	Social approach with top 15 similar users	Amplified approach with top 10 similar users
Unidirectional readership ISN	Amplified approach with top 20 similar users	Amplified approach with top 20 similar users
Co-readership ISN	Combined approach with top 20 similar users	Combined approach with top 20 similar users
Tag-based ISN	Amplified approach with top 20 similar users	Amplified approach with top 20 similar users

6.2 The Role of Different Social Distances on Recommendation

First, we used Levene's test for equality of variance to check the assumption of homogeneity; if this assumption was violated ($p < 0.05$), then we used Welch's ANOVA rather than one-way ANOVA. We further explored any statistically significant differences between groups using post hoc tests, either Tukey or Games-Howell depending on whether the homogeneity assumption was met.

We found statistically significant differences in prediction accuracy (i.e., F1 measure at different ranks) and diversity at different social distances (i.e., direct, one hop or two hops) in the reciprocal and unidirectional readership ISNs (see Table 6-2 and Table 6-3). Therefore, we applied post hoc tests (Tukey's test when Levene's test was not significant and Games-Howell when it was) and found that when the social relations were reciprocal, the recommendations based on direct user relationships had higher prediction accuracy than when the social relations were indirect. The difference was not statistically significant when only two research papers were recommended (F1@2), but the differences were significant otherwise. When more research papers were recommended (i.e., N = 10, 15 or 20), there were no statistically significant differences in prediction accuracy between one hop and two, but accuracy for N = 2 or N = 5 was significantly higher at one hop's distance than with two hops. Post hoc tests for diversity at different ranks showed that using two hops provided more diverse recommendation lists than did using direct or one-hop social relations, and this difference was statistically significant.

Table 6-2: Comparison of different measures considering three distances for reciprocal readership ISN

Measure	Direct	One hop	Two hops	Levene test	One-way ANOVA, p=.005	Welch's ANOVA, p=.005
F1@2	.1331±.0078	.1069±.0328	.0250±.0184	p=.029		**(2,6.353)=67.950**
F1@5	.1638±.0086	.0954±.0364	.0392±.0264	p=.281	***F*(2,12)=26.917**	
F1@10	.1462±.0086	.0802±.0291	.0479±.0277	p=.143	***F*(2,12)=22.807**	
F1@15	.1195±.0055	.0626±.0264	.0375±.0176	p=.057	***F*(2,12)=25.505**	
F1@20	.1022±.0051	.0503±.0196	.0336±.0129	p=.086	***F*(2,12)=33.293**	
Div@2	.0121±.0022	.0153±.0058	.0515±.0098	p=.345	***F*(2,12)=53.470**	
Div@5	.0095±.0010	.0087±.0042	.0546±.0121	p=.005		**(2,5.657)=30.845**
Div@10	.0074±.0005	.0078±.0021	.0490±.0084	p=.005		**(2,5.527)=54.543**
Div@15	.0061±.0003	.0071±.0043	.0451±.0067	p=.013		**(2,5.360)=76.242**
Div@20	.0049±.0001	.0074±.0039	.0427±.0041	p=.007		**(2,5.348)=189.226**

The post hoc tests for the *unidirectional readership ISN* showed consistent results for prediction accuracy and diversity at different social distances. Both accuracy and diversity with two-hop social relations were statistically significant higher than the results of one hop and direct relations, and there were also significant differences between the results for one-hop and direct relations; at one hop, the results were significantly higher.

Table 6-3: Comparison of different measures considering three distances for unidirectional readership ISN

Measure	Direct	One hop	Two hops	Levene test	One-way ANOVA at p<.0005
F1@2	.0351±.0023	.0403±.0400	.0440±.0440	p=.702	***F*(2,12)=23.014**
F1@5	.0473±.0015	.0513±.0013	.0553±.0015	p=.908	***F*(2,12)=38.966**
F1@10	.0514±.0010	.0528±.0010	.0553±.0011	p=.684	***F*(2,12)=20.468**
F1@15	.0504±.0010	.0510±.0010	.0527±.0013	p=.081	***F*(2,12)=7.160**
F1@20	.0480±.0007	.0486±.0010	.0499±.0010	p=.557	***F*(2,12)=7.526**
Div@2	.0065±.0001	.0093±.0010	.0108±.0004	p=.019	***F*(2,12)=127..312**
Div@5	.0048±.0001	.0067±.0003	.0077±.0002	p=.292	***F*(2,12)=275.307**
Div@10	.0036±.0006	.0049±.0008	.0056±.0001	p=.283	***F*(2,12)=608.210**
Div@15	.0030±.0004	.0040±.0006	.0046±.0008	p=.320	***F*(2,12)=805.862**
Div@20	.0026±.0004	.0035±.0005	.0040±.0006	p=.445	***F*(2,12)=901.592**

Table 6-4 illustrates that the one-way ANOVA and Welch's ANOVA results for the co-readership ISN showed statistically significant differences in some of the results. For the significant results (F1@2, F1@5), we used post hoc tests that showed that the recommendation performances for direct and one-hop social relations were significantly higher than performance when the two-hop distance was used. However, there were no statistically significant differences between F1 values when the relationship between two connected users was direct or indirect with

1 hop social distance. The diversity of the recommendation lists at different social relation distances was significant only when 5 or 10 research papers were recommended, and the diversity with indirect relations was significantly higher than that with directly connected users. However, the list's diversity was the same at either one or two hops in social relation distance.

Table 6-4: Comparison of different measures considering three distances for co-readership ISN

Measure	Direct	One hop	Two hops	Levene test	One-way ANOVA	Welch's ANOVA , p=.005
F1@2	.03518±.00048	.0355±.0007	.0343±.0008	p=.753	***F*(2,12)=4.152, p=0.43**	
F1@5	.04540±.00051	.0455±.0006	.0444±.0005	p=.876	***F*(2,12)=6.368,p=.013**	
F1@10	.04704±.00022	.0468±.0007	.0465±.0008	p=.012		(2,6.142)=.977, p=.428
F1@15	.04499±.00022	.0446±.0005	.0445±.0006	p=.034		(2,6.549)=1.687,p=.257
F1@20	.04232±.00025	.0421±.0005	.0418±.0005	p=.230	*F*(2,12)=1.584,p=.245	
Div@2	.00256±.00004	.0026±.00006	.0026±.00006	p=.655	*F*(2,12)=2.378,p=.135	
Div@5	.00186±.00003	.00191±.00001	.00191±.0001	p=.141	***F*(2,12)=6.466,p=.012**	
Div@10	.00140±.00002	.00143±.00002	.00143±.00001	p=.202	***F*(2,12)=5.883,p=.017**	
Div@15	.00117±.00002	.00119±00001	.00119±.00001	p=.537	*F*(2,12)=3.642,p=.058	
Div@20	.00102±.00001	.00104±00001	.00104±.00001	p=.697	*F*(2,12)=3.350,p=.070	

For the tag-based ISN, Table 6-5 shows that there were no statistically significant results at different social relation distances, and thus, there was no need for follow-up tests. The results show that social distance had no effect on prediction accuracy or recommendation diversity when the tag-based ISN was the recommendation information source.

Table 6-5: Comparison of different measures considering three distances for tag-based ISN

Measure	Direct	One hop	Two hops	Levene test	One-way ANOVA	Welch's ANOVA, p=.005
F1@2	.0326±.0012	.0327±.0012	.0324±.0010	p=.590	*F*(2,12)=.089,p=.915	
F1@5	.0425±.0010	.0430±.0012	.0425±.0008	p=.440	*F*(2,12)=.394,p=.683	
F1@10	.0443±.0010	.0450±.0010	.0446±.0010	p=.889	*F*(2,12)=1.146,p=.350	
F1@15	.0429±.0003	.0432±.0006	.0429±.0004	p=.144	*F*(2,12)=.891,p=.436	
F1@20	.0407±.0003	.0409±.0005	.0409±.0002	p=.123	*F*(2,12)=.542,p=.595	
Div@2	.0029±.0001	.0029±.0004	.0029±.0007	p=.113	*F*(2,12)=.091,p=.913	
Div@5	.0021±.0002	.0021±.0003	.0021±.0001	p=.131	*F*(2,12)=.043,p=.958	
Div@10	.0016±.0001	.0016±.0002	.0016±.0003	p=.042		(2,6.491)=.122, p=.887
Div@15	.0013±.0001	.0013±.0001	.0013±.0002	p=.036		(2,36.591)=.625, p=.564
Div@20	.0011±.0001	.0011±.0001	.0011±.0001	p=.059	*F*(2,12)=.131,p=.878	

6.3 Exploring the Trade-off between Accuracy and Diversity

Research has shown that nonperformance measures such as diversity of recommendation lists increases at the expense of prediction accuracy [136]. In this section, we explore the gains in diversity versus the losses in prediction accuracy at different social relation distances between users; specifically, we hypothesized that diversity would increase and prediction accuracy would decrease at greater social distance levels. We compared the F1 measure as the prediction measure with the intra-list diversity measure for different social distances (direct, one hop and two hops) for each ISN; Figure 6-1 shows the accuracy-diversity curves for the three different social distances. From Figure 6-1 and Table 6-6, we can observe that the loss in accuracy and the gain in diversity at one hop in the *reciprocal readership ISN* were lower than those at two hops; recommendation lists based on two hops were more diverse than those based on direct social relations between users. However, as the one-way ANOVAs showed previously, the difference in diversity between direct and one-hop relations was not significant whereas the F1 difference was. We can conclude that using data based on one-hop social network relations is more effective than using bookmarked papers from users at a two-hop distance in that the total utilities were -5.2707 and -5.5804, respectively.

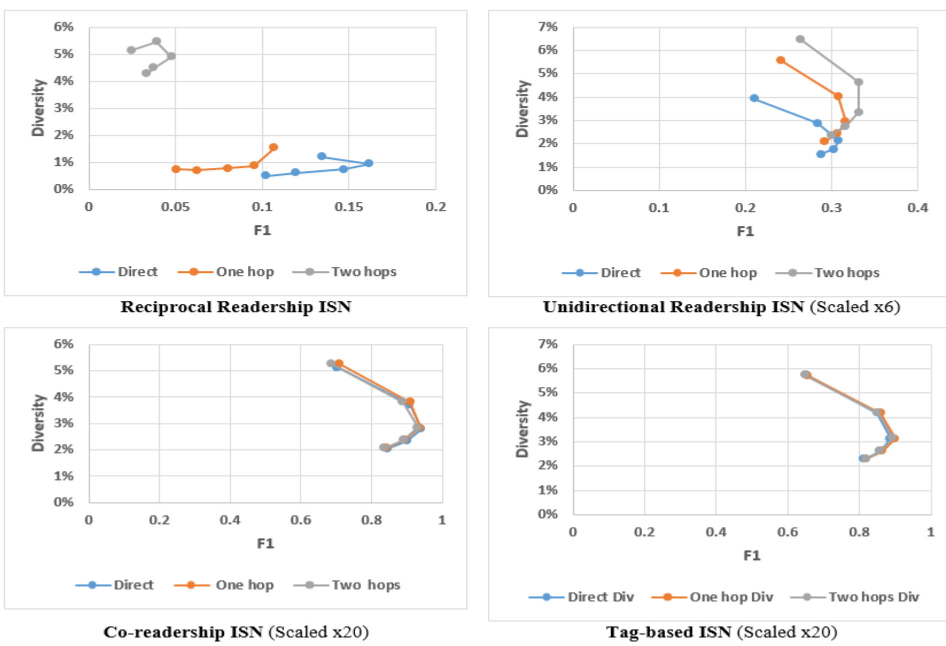


Figure 6-1: Comparison between different accuracy-diversity curves according to different social distances for each ISN.

The results for the unidirectional readership ISN were interesting. As shown in Figure 6-1 and Table 6-6, both accuracy and diversity increased with increased social distance, and the results were best at the two-hop distance. Thus, it is not necessary that the diversity always increase with the expense of prediction accuracy.

Table 6-6: Loss (-) and gain (+) in accuracy and diversity when considering different social relation distances in each ISN

ISN	One hop distance		Two hops distance	
	F1	Diversity	F1	Diversity
Reciprocal readership	-5.3927 %	+0.1220%	-9.6351%	+4.0547%
Unidirectional readership	+0.2361%	+0.1604%	+0.5004%	+0.2441%
Co-readership ISN	-0.0073	+0.0035%	-0.0664%	+0.0035%
Tag-based ISN	+0.0334%	+0.0003%	+0.0065%	+0.0004%

Figure 6-1 shows that the prediction-diversity curve for the *co-readership ISN* at one hop's distance was slightly dominant over the other two curves. Table 6-6 shows that accuracy decreased and diversity increased at greater social distance; the total differences in accuracy and diversity between the direct and one-hop relations and the direct and two-hop relations were -0.0038 and -0.0629, respectively. Therefore, using data from users at a one-hop distance can increase diversity with minimal loss of accuracy.

The prediction-diversity curves for the tag-based ISN in Figure 6-1 show that the three curves are highly overlapped, which reflects no significant difference, possibly because there were insufficient tags in the dataset to improve the results.

6.4 Summary

In objective of this chapter is to study the effect of considering different social relations between users on recommendation. We studied the role of the social distance on prediction accuracy and on the diversity of the recommended list. First, the same three recommendation approaches that were applied on the data of network of users who are directly connected were also applied on the networks of users who are connected distantly using one intermediate user (one hop) or two intermediate users, and the best performed approach was chosen from each ISN with one hop distance or two hops distance. Then, the prediction accuracy and the diversity of recommendation are compared with taking into account the social distance for each ISN. Lastly, the trade-off between the prediction accuracy and the diversity are compared for each network by drawing the F1-diversity curves for direct, one hop, and two hops distances for each ISN.

CHAPTER 7
SUMMARY, LIMITATIONS AND FUTURE WORK

This dissertation aims to alleviate some of the problems with the collaborative filtering approach. A comprehensive literature survey was conducted that revealed a lack of research in the area of recommender systems using multi-criteria ratings in general and in recommending research papers specifically. Although similarities exist between multi-criteria recommender systems, context-aware recommender systems and content-based recommender systems, multi-criteria recommender systems (MCRS) consider the subjective opinions of the user about the rating criteria while context-aware and content-based recommenders take into account the objective features of the items. Systems using multi-criteria ratings reduce ambiguity and allow users to provide more details about the quality of items, in comparisons with systems using single rating. In addition, researchers have rarely found systems with numeric ratings for recommending research papers. We also found very few studies examining the social relationships among users in recommending research papers, and most of this research uses relationships initiated by users, which indicate that recommendations can be provided to a small number of users.

7.1 Summary of Findings

In this dissertation, we proposed two ideas aiming to mitigate some of CF's drawbacks such as data sparsity, the cold start problem, and overall rating bias. First, we proposed using multi-criteria rating instead of one overall rating as used in conventional CF. Second, we exploited the wealth of user data stored by social bookmarking tools (e.g., CiteULike, Mendeley) to construct new social relationships among users that can connect more users. The findings related to these two proposals are discussed below.

7.1.1 Summary of findings for multi-criteria rating studies

For this part of the dissertation, we wanted to answer research question RQ1, "How do users perceive the multi-criteria rating recommendations?" To answer this question, we conducted a focus group study. This approach is suited to exploratory studies, and it enabled us to gather users' opinions. We followed this with an online questionnaire (69 participants) to confirm the findings from the focus group study and to compare the results of participants from different disciplines (please refer to sections 3.1 and 3.2). RQ1 divided into four specific research questions.

The first specific question (RQ1.1) was “What are the most important rating criteria for evaluating a research paper?” To answer this question, participants were asked to come up with every criterion that could be used to evaluate a research paper. After this brainstorming session, they were asked to list all the criteria in descending order starting from the most important criterion. They chose the following criteria: clarity, technical clarity, and relevance to their research interest. Interestingly, the participants thought that the paper’s relevance to their interest must be rated even though the system evaluates this criterion by matching the user’s interest with the research paper’s features. Their justification was that some users have multidisciplinary research fields, and it is important for them to confirm the relevance of research papers to their research interests.

The second specific research question (RQ1.2) was about users’ preferences for overall ratings versus multi-criteria ratings. We found that participants prefer to have both overall ratings and a short list of multi-criteria ratings. Their justification was that overall ratings enable users to state their overall impressions about research papers, while multi-criteria ratings enable them to evaluate different aspects of any given paper’s quality. Then, the recommender system finds papers that fulfill their preferences according to the rating criteria.

Most participants in the quantitative study and all participants in the focus group study showed strong support for enabling the users of MCRS to control the importance weight of each rating criterion. This allowed us to answer affirmatively RQ1.3, “Do users prefer to have control over the importance weights of multi-criteria ratings during the recommendation process?” Users prefer to show their preferences and be involved in the recommendation process. In our study, participants were excited about the idea of being involved in the recommendation process, and they suggested some interface designs to capture users’ weights of ratings criteria by using sliders or by ranking the criteria. They pointed out some interesting benefits of being able to change the importance weight of rating criteria, since the relative importance of the rating criteria differs from user to user, and for a given user in time. Thus, if the user’s objectives change over time, she can cope with this change by increasing the weight of more important criteria or by decreasing the weight of less important criteria.

RQ1.4 determines whether the criteria are domain-dependent. In the questionnaire, we asked participants from two different disciplines to rank the rating criteria. We found that all participants chose the same first two rating criteria even though they were from completely different disciplines. However, participants from Computer Science chose technical clarity and clarity as the third and fourth criteria, respectively, while participants while participants from pharmacy and nutrition ranked clarity higher than technical clarity. However, the total scores for the clarity and technical clarity were almost the same. Thus, we can conclude that there are no significant differences in the ranking of different criteria by users from different disciplines, and the rating criteria are domain-independent.

7.1.2 Summary of findings for the implicit social networks studies

In the second part of this dissertation focusing on social relationships among users, we constructed three social networks based on the similarities between users that shared certain features. Then we studied similarities among connected users in each of the proposed social networks and answered RQ2 to RQ4 in sections 4.6 and 4.7. RQ2 investigated which implicit social network connects the most similar users. We found that the readership ISN connects users with the highest similarity compared to the other ISNs, and the similarity value is higher when social relationships are reciprocal. Thus, users share similar interests with the authors of the papers that they bookmarked in their libraries. Then, the co-readership ISN; then the tag-based ISN connects users with lowest interest similarity value. The results using different similarity measures of bookmarking behaviour are all consistent. That is to say, all the measures gave the same exact results. For this reason, we used the LogLikelihood measure, a relative measure for unary rating, for the rest of our experiments.

We also tested the effect of social relation distance on the similarity between two connected users of each ISN (RQ3). We found that the highest similarity value was achieved when the social relation is direct, and the interest similarity decreases when the social relation distance increases. We tested this using three different social distances: direct, indirect using one intermediate user between two connected users, and indirect using two intermediate users between two connected users. This means users who are socially connected indirectly still have some degree of similarity, which has useful applications, especially for the cold start problem. The interest similarity among connected users of proposed ISNs was compared to the interest similarity of connected users of two known explicit social networks. This allowed us to answer affirmatively RQ4 ”Does the interest similarity between users who are implicitly socially connected compare with that between users who are explicitly connected?”. The readership ISN connects users with more similarities than connected users in co-authorship and friendship explicit social networks.

The rest of the dissertation work tested recommendations using performance measures (i.e., prediction accuracy) and nonperformance measures, such as user coverage and the diversity of recommendation list, when data from only ISNs were used or when data from both ISNs and explicit SN were combined. The broad question RQ5 is “What is the effect of using implicit social networks on improving recommendations?”

This broad question is divided into four specific questions. First, we wanted to compare recommendations using only data from social friends (i.e., connections) with recommendations using data from both social friends and anonymous peers (i.e., using CF). Thus, we compared three different recommendation approaches: social, combined, and amplified. Our goal was to find which one works best with each ISN (RQ5.1, section 5.3.1). In the social approach, the recommendation was produced after collecting research papers from only the user’s social friends and by substituting a similarity score that is usually used in CF with the weighted strength of the relationships among socially connected users. While this combined approach gives data from social friends of the target user the same weight as data from anonymous peers, the amplified approach gives more weight to the papers bookmarked by the user’s social friends than those papers bookmarked by anonymous peers. We applied all of the three recommendation approaches to the data from each ISN. We found for the ISN that has the strongest social relationships among users (i.e., reciprocal readership ISN), the social approach worked best. For the weakest social relationships (i.e., tag-based ISN), the amplified approach worked best. The combined approach worked best for the unidirectional readership and co-readership ISNs. In addition, the results of different recommendation approaches were compared to CF, and we found that the approaches for each ISN that use the social data outperform or at least perform similar to those of CF.

Then we compared the recommendation produced by each ISN to answer RQ5.2 “Which ISN produces the highest prediction accuracy?” We found that the highest prediction accuracy using precision and recall metrics was achieved when the reciprocal readership ISN was used as the information source for recommendation, followed by the unidirectional readership and co-readership ISN with no statistically significant differences in their prediction accuracy values. The tag-based ISN which was the least effective.

However, taking only prediction accuracy into account is not enough in evaluating recommender systems. The literature shows a trade-off between prediction accuracy and user coverage. This trade-off was clear between the prediction accuracy and user coverage when different ISNs were used. For example, the highest prediction accuracy was achieved when reciprocal readership was used, but the user coverage was the lowest of all ISNs. For this reason, we fused recommendations from implicit and explicit social networks to strike a balance between prediction accuracy and user coverage of recommendations and to answer RQ5.3 (sections 5.4.2 and 5.4.3). We tested fusing recommendations from either friendship social networks or co-authorship social networks with every ISN: Fusing recommendations from implicit and explicit social networks enhances both prediction accuracy and user coverage in the case of friendship social networks. User coverage improved in the case of co-authorship social networks without significant differences in the prediction accuracy.

The last experiment was to answer RQ5.4 “What is the effect of social relationship distance on prediction accuracy and the diversity of recommendation lists?” (Results discussed in section 6.3). The accuracy-diversity curves of different social distances were drawn for each ISN. We found that a one-hop distance curve struck the best balance between accuracy and diversity for the reciprocal readership and co-readership ISNs. The accuracy-diversity curves for the tag-based ISN were almost the same. However, the curves for the unidirectional readership ISN showed that the two-hop distance relation is the dominant curve.

7.2 Contributions

This dissertation contributes to two areas: multi-criteria rating recommender systems and social network-based recommender systems, with a focus on recommending research papers. It also contributes generally to the fields of user modelling, personalization, and adaptation. Because the items under consideration are research papers, our hope is that this dissertation will have an impact on lifelong learning. Some parts of this dissertation suggest ideas for enhancing user interfaces for recommender systems, which contributes to the field of human-computer interaction (HCI). Specifically, the following contributions have been made:

- *User requirement analysis through qualitative methods*: Qualitative research methods, such as interviews or focus groups, are well-known for collecting participants' opinions and insights about a given topic. However, qualitative studies are rarely used to collect user requirements before designing recommender systems, which results in systems that do not fulfill the user's needs. As a consequence, users might not adopt the recommender system. In this dissertation, we explored users' opinions regarding a recommender system that uses multi-criteria ratings by conducting a focus group study. This allowed for direct interactions with and among the participants and enabled an exploration and redefinition of the scope of ideas.
- *Development of alternative methods to connect users socially:* In social bookmarking tools such as Mendeley and CiteULike, the focus is on users' finding papers to bookmark. Users pay attention to bookmarking more than connecting with other users. Thus, there are only few explicit social relationships. For this reason, we used the available data about bookmarked papers to find other kinds of social relationships that connect more users than explicit social relationships.
- *Discovery of the trade–offs between prediction accuracy and user coverage measures*: While most studies focus on increasing prediction accuracy by developing different recommendation approaches, we argue that prediction accuracy must be studied in conjunction with other measures. Increasing prediction accuracy and providing recommendations to only a few users is not better than maintaining the same prediction accuracy but increasing the number of users to which the recommender system can provide recommendations. We studied the effects of using different implicit social networks on prediction accuracy and user coverage.
- *Discovery of the trade-offs between prediction accuracy and diversity measures:* Researchers have found that the more accurate the prediction, the more similar the recommended items. We tested the diversity of the recommendation list when social relation propagation through social networks is used, and when different types of implicit social networks are used. To the best of our knowledge, this is the first work that studies the trade-off between prediction accuracy and diversity in recommending research papers when different social relation distances are considered.

- *Development of a hybrid approach that fuses recommendations from different social network types:* We proposed fusing recommendations that use data from different social network types (implicit and explicit social networks). We tested fusing recommendations from different combinations of two known explicit social networks (co-authorship and friendship) with the proposed ISNs. We considered prediction accuracy and user coverage. Most hybrid approaches look at combining the results of different recommending approaches, such as CF, with CBF. However, we fused different recommendations from different sources of information (i.e. different social data).

7.3 Implications for Designing Recommender Systems to Recommend Research Papers

The findings of this dissertation support the following implications of designing recommender systems for research papers:

- The research shows that users do not perceive how recommendations are produced. Thus, designers should explain recommendations by adding some comments that inform users which data were used to produce each recommendation or by including some visual tools.
- We found that participants were excited about allowing users to change the weight of different rating criteria, and they suggested some user interface designs (please refer to Table 3-3). Designers should carefully study how users can provide the system with their input and conduct a usability study of the designed interface. In addition, designers of algorithms should also study the relationships among different rating criteria to allow users to control the weights effectively.
- Participants in our studies were interested in both overall ratings and multi-criteria ratings. We think that including both kinds of ratings might be useful for two reasons. First, the overall rating can be used to provide recommendations using the conventional CF approach. It is also good for users who do not want to provide many ratings. Second, the overall rating can be used to infer the relationship between the multi-criteria ratings and the overall rating and help enrich the user's profile once the designers of recommendation algorithms find that relationship.
- We have shown that participants were insistent that the relevance of a research paper should be one of the rating criteria. We argue that rating the relevance of a paper is beneficial because the user's research interests change over time. Thus, using the

relevancy rating along with the date of rating can help in better modelling the user's interests.

- We found that qualitative studies are useful to understand how users think and what they need from the system. We encourage designers to conduct qualitative studies prior to design recommender system to confirm or negate their ideas.
- One of our findings confirms that fusing recommendations from different kinds of social networks improves prediction accuracy and user coverage. Thus, designers of recommendation algorithms should take into account the different explicit and implicit social relationships that users might have according to the data about user activities mediated through objects (e.g. co-bookmarking papers, co-authoring papers) in recommendations. It is also worth studying user involvement in choosing the source of the recommendation and its weight in the recommendation fusing process by allowing the user to change the weight of different recommendation sources. For example, users could increase the weight of recommendations from their own friends.
- Propagation through social networks is beneficial to increase the diversity of the recommendations. Designers of social recommender systems should include recommendations from distant social friends as well as from direct friends to create a mix of papers with different diversity scores.

7.4 Limitations of the Study and Potential Future Directions

This work described in this dissertation has several limitations. The first part of the dissertation focused on users' opinions regarding the most important rating criteria for designing multi-criteria rating recommenders. Most users who participated in either the focus groups or the questionnaires were students or young researchers (postdoctoral fellows). Young researchers and experienced researchers might think differently and consider different criteria in deciding which papers to read. For example, young researchers might need more review papers to help them understand their research areas. Furthermore, participants in these studies are only from two research domains: computer science and pharmacy and nutrition. The objective was to test whether participants from these two different areas held different opinions about the importance of specific rating criteria. However, both fields are considered science domains. The criteria might be different if we consider other domains such as humanities or social sciences. Collecting more data from different user groups (e.g., undergraduate students or professors) and from different disciplines is necessary to generalize our findings. In addition, we could not continue working on the area of using multi-criteria rating to recommend research papers because of the absence of suitable dataset as explained in section 3.3. In the future, we plan to gather multi-criteria ratings from real users and test the effect on the prediction accuracy and the user satisfaction of the recommendation of the deployed system.

Another limitation of this work is that we conducted offline experiments to evaluate the recommendations that use ISNs as sources of information. The objective was to extensively study the prediction accuracy of the recommendations produced using ISNs and compare it to the prediction accuracy of hybrid approaches using data from ISNs and explicit social networks along with studying the effects of social distance on prediction accuracy. Although we studied the relationship between prediction accuracy measures and nonperformance measures in this dissertation, some other measures cannot be studied using offline experiments, and user studies must be conducted to evaluate other aspects of recommendations such as user satisfaction. To fill this gap, user studies could be done to compare the results from offline experiments with the results from user studies. Another limitation is that only one dataset was used in all of the offline experiments. To generalize our findings, testing ISNs using other datasets is recommended.

Other potential future research directions include recommendations for users to connect to one another. Recommendations based on proposed ISNs use data from users that the target user does not know. Therefore, users who are similar to the target user using ISNs either connected directly or indirectly can be good for people recommendation.

Our work can be generalized to other domains that use social bookmarking tools. Once the system can save bookmarks of one kind of information, we can infer social relationships among users with the available data. In this dissertation, we developed three implicit social networks, but more data is available and there are more potential implicit networks that can be generated and exploited, using bookmarked object metadata, user-generated data, or a combination of both.

In this work, we used the same weights for all users to combine recommendations from different resources, which could limit the personalization capabilities. In the future, we want to test using dynamic weights that are based on each user's features such as social network features (e.g. number of incoming/outgoing social relations). In addition, we want to test the effect of giving the user the control on the fusing weights for explicit and implicit social network-based recommendations.

7.5 Conclusion

This dissertation explored two areas in recommending research papers: multi-criteria rating systems that use the subjective opinions of users regarding different aspects of research papers and social network-based recommender systems that use social relationships among users as extra input. We conducted qualitative and quantitative studies to explore users' perceptions of multi-criteria rating systems for recommending research papers and their preference for being involved in the recommendation process. We also conducted a series of offline experiments to test the feasibility of implicit social networks as a good information source for recommendations. We tested a limited number of features to infer social connections among users that are based on metadata or user-generated data of the published or bookmarked papers in their libraries. However, we paid more attention to performance and nonperformance measures and the trade-off between them.

We found that participants were excited about using multi-criteria systems, and they were strongly supportive of the idea of enabling users to be part of the recommendation process by changing the importance weight for each rating criterion. We also found that users rank different rating criteria differently, which also supports the idea that designers of recommender systems should allow different weights for criteria.

We also found that user data accumulated by social bookmarking systems different can be used to generate implicit social networks connecting users which can increase prediction accuracy and user coverage in finding users with similar interests, and fusing recommendations from explicit social networks with implicit social networks strikes a good balance between prediction accuracy and user coverage.

We conclude that we see promise in both multi-criteria rating systems and implicit social networks for recommending research papers. Future research will investigate how such recommender systems could be used to recommend other items and services as well and how users can be empowered to understand and control the recommendation process.

REFERENCES

[1] C. Basu, H. Hirsh, W. W. Cohen, and C. Nevill-Manning, “Technical paper recommendation: a study in combining multiple information sources,” *J Artif Int Res*, vol. 14, no. 1, pp. 231–252, May 2001.

[2] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl, “On the recommending of citations for research papers,” in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New York, NY, USA, 2002, pp. 116–125.

[3] S. Jung, K. Harris, J. Webster, and J. L. Herlocker, “SERF: integrating human recommendations with search,” in *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, New York, NY, USA, 2004, pp. 571–580.

[4] S. E. Middleton, D. C. De Roure, and N. R. Shadbolt, “Capturing knowledge of user preferences: ontologies in recommender systems,” in *Proceedings of the 1st international conference on Knowledge capture*, New York, NY, USA, 2001, pp. 100–107.

[5] A. Naak, H. Hage, and E. Aïmeur, “A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyres,” vol. 26, G. Babin, P. Kropf, and M. Weiss, Eds. Springer Berlin Heidelberg, 2009, pp. 25–39.

[6] G. Adomavicius and Y. Kwon, “New Recommendation Techniques for Multicriteria Rating Systems,” *IEEE Intell. Syst.*, vol. 22, no. 3, pp. 48–55, May 2007.

[7] N. Tintarev and J. Masthoff, “Effective Explanations of Recommendations: User-centered Design,” in *Proceedings of the 2007 ACM Conference on Recommender Systems*, New York, NY, USA, 2007, pp. 153–156.

[8] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a Feather: Homophily in Social Networks,” *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.

[9] P. V. Marsden and N. E. Friedkin, “Network Studies of Social Influence,” *Sociol. Methods Res.*, vol. 22, no. 1, pp. 127–151, Aug. 1993.

[10] D. Lee, “Personalized Recommendations Based On Users’ Information-Centered Social Networks,” University of Pittsburgh, 2013.

[11] D. H. Lee and P. Brusilovsky, “Interest Similarity of Group Members: the Case study of Citeulike,” presented at the WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC: US, 2010.

[12] D. H. Lee and P. Brusilovsky, “Improving Recommendations Using WatchingNetworks in a Social Tagging System,” in *Proceedings of the 2011 iConference*, New York, NY, USA, 2011, pp. 33–39.

[13] G. Shani and A. Gunawardana, “Evaluating Recommendation Systems,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 257–297.

[14] J. L. Herlocker, J. A. Konstan, L. G. Terveen, John, and T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Trans. Inf. Syst.*, vol. 22, pp. 5–53, 2004.

[15] J. Tang, X. Hu, and H. Liu, “Social recommendation: a review,” *Soc. Netw. Anal. Min.*, vol. 3, no. 4, pp. 1113–1133, Nov. 2013.

[16] R. Burke, “Hybrid Recommender Systems: Survey and Experiments,” *User Model. User-Adapt. Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[17] K. Lang, “NewsWeeder: Learning to Filter Netnews,” in *in Proceedings of the 12th International Machine Learning Conference (ML95*, 1995.

[18] J. B. Schafer, J. Konstan, and J. Riedi, “Recommender systems in e-commerce,” in *Proceedings of the 1st ACM conference on Electronic commerce*, New York, NY, USA, 1999, pp. 158–166.

[19] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information tapestry,” *Commun ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992.

[20] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, New York, NY, USA, 1994, pp. 175–186.

[21] J. Beel and C. Hentschel, “Scienstein – A Research Paper Recommender System,” in *Proceedings of the International Conference on Emerging Trends in Computing (ICETiC’09)*, Virudhunagar (India), 2009, pp. 309–315.

[22] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, “Enhancing digital libraries with TechLens+,” in *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, 2004, pp. 228–236.

[23] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowl. DATA Eng.*, vol. 17, no. 6, pp. 734–749, 2005.

[24] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, “Probabilistic memory-based collaborative filtering,” *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 1, pp. 56–69, Jan. 2004.

[25] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, “Combining Content-Based and Collaborative Filters in an Online Newspaper,” in *Proceedings of the ACM SIGIR ’99 Workshop on Recommender Systems: Algorithms and Evaluation*, Berkeley, California, 1999.

[26] C. Dellarocas, “Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior,” in *Proceedings of the 2Nd ACM Conference on Electronic Commerce*, New York, NY, USA, 2000, pp. 150–157.

[27] 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, USA, 2004, pp. 393–402.

[28] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, San Francisco, CA, USA, 1998, pp. 43–52.

[29] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering,” in *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 1999, pp. 230–237.

[30] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based Collaborative Filtering Recommendation Algorithms,” in *Proceedings of the 10th International Conference on World Wide Web*, New York, NY, USA, 2001, pp. 285–295.

[31] B. Krulwich, “LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data.,” *AI Mag.*, vol. 18, no. 2, pp. 37–45, 1997.

[32] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, “Eigentaste: A Constant Time Collaborative Filtering Algorithm,” *Inf Retr*, vol. 4, no. 2, pp. 133–151, Jul. 2001.

[33] T. Hofmann, “Collaborative filtering via gaussian probabilistic latent semantic analysis,” in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, New York, NY, USA, 2003, pp. 259–266.

[34] B. Sarwar, G. Karypis, J. Konstan, and J. Rield, *Analysis of Recommendation Algorithms for E-Commerce*. 2000.

[35] H. Ma, H. Yang, M. R. Lyu, and I. King, *SoRec: Social Recommendation Using Probabilistic Matrix Factorization*. 2008.

[36] D. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, “Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach,” in *In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 2000, pp. 473–480.

[37] D. Billsus and M. J. Pazzani, “User Modeling for Adaptive News Access,” *User Model. User-Adapt. Interact.*, vol. 10, no. 2–3, pp. 147–180, Feb. 2000.

[38] A. M. A. Wasfi, “Collecting User Access Patterns for Building User Profiles and Collaborative Filtering,” in *In Proceedings of the 1999 International Conference on Intelligent User Interfaces*, 1999, pp. 57–64.

[39] R. J. Mooney and L. Roy, “Content-based book recommending using learning for text categorization,” in *Proceedings of the fifth ACM conference on Digital libraries*, New York, NY, USA, 2000, pp. 195–204.

[40] C. Basu, H. Hirsh, and W. Cohen, “Recommendation as classification: using social and content-based information in recommendation,” in *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, Menlo Park, CA, USA, 1998, pp. 714–720.

[41] M. Balabanović, “An adaptive Web page recommendation service,” in *Proceedings of the first international conference on Autonomous agents*, New York, NY, USA, 1997, pp. 378–385.

[42] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach," *ACM Trans. Inf. Syst.*, vol. 23, pp. 103–145, 2005.

[43] G. Adomavicius, N. Manouselis, and Y. Kwon, "Multi-Criteria Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 769–803.

[44] W.-P. Lee, C.-H. Liu, and C.-C. Lu, "Intelligent agent-based systems for personalized recommendations in Internet commerce," *Expert Syst. Appl.*, vol. 22, no. 4, pp. 275–284, May 2002.

[45] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 217–253.

[46] T. Y. Tang and G. McCalla, "A Multidimensional Paper Recommender: Experiments and Evaluations," *IEEE Internet Comput.*, vol. 13, no. 4, pp. 34–41, Aug. 2009.

[47] N. Sahoo, R. Krishnan, G. Duncan, and J. P. Callan, "Collaborative filtering with multi-component rating for recommender systems," in *Proceedings of the sixteenth workshop on information technologies and systems*, 2006.

[48] L. Si and R. Jin, "Flexible Mixture Model for Collaborative Filtering," in *In Proc. of ICML*, 2003, pp. 704–711.

[49] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of Dimensionality Reduction in Recommender System – A Case Study," in *IN ACM WEBKDD WORKSHOP*, 2000.

[50] W.-P. Lee, C.-H. Liu, and C.-C. Lu, "Intelligent agent-based systems for personalized recommendations in Internet commerce," *Expert Syst. Appl.*, vol. 22, no. 4, pp. 275–284, May 2002.

[51] H. Kautz, B. Selman, and M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering," *Commun ACM*, vol. 40, no. 3, pp. 63–65, Mar. 1997.

[52] J. Golbeck, "Generating predictive movie recommendations from trust in social networks," in *Proceedings of the 4th international conference on Trust Management*, Berlin, Heidelberg, 2006, pp. 93–104.

[53] 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*, New York, NY, USA, 2008, pp. 931–940.

[54] B. Sigurbjörnsson and R. van Zwol, "Flickr Tag Recommendation Based on Collective Knowledge," in *Proceedings of the 17th International Conference on World Wide Web*, New York, NY, USA, 2008, pp. 327–336.

[55] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *Proceedings of the fourth ACM conference on Recommender systems*, New York, NY, USA, 2010, pp. 199–206.

[56] 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*, R. Meersman and Z. Tari, Eds. Springer Berlin Heidelberg, 2004, pp. 492–508.

[57] A. Bellogin, I. Cantador, and P. Castells, "A Study of Heterogeneity in Recommendations for a Social Music Service," in *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, New York, NY, USA, 2010, pp. 1–8.

[58] F. Liu and H. J. Lee, "Use of social network information to enhance collaborative filtering performance," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 4772–4778, Jul. 2010.

[59] J. He and W. W. Chu, "A Social Network-Based Recommender," in *Data Mining for Social Network Data*, N. Memon, J. J. Xu, D. L. Hicks, and H. Chen, Eds. Springer US, 2010, pp. 47–74.

[60] S. Bourke, K. McCarthy, and B. Smyth, "Power to the people: exploring neighbourhood formations in social recommender system.," in *ResearchGate*, 2011, pp. 337–340.

[61] and R. Sinha and K. Swearingen, "Comparing Recommendations Made by Online Systems and Friends," in *In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*, 2001.

[62] G. Groh and C. Ehmig, "Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering," in *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, New York, NY, USA, 2007, pp. 127–136.

[63] 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 - Volume 1*, Swinton, UK, UK, 2007, pp. 77–86.

[64] Q. Yuan, S. Zhao, L. Chen, Y. Liu, S. Ding, X. Zhang, and W. Zheng, "Augmenting collaborative recommender by fusing explicit social relationships," in *Workshop on Recommender Systems and the Social Web, Recsys 2009*, 2009.

[65] 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*, New York, NY, USA, 2009, pp. 195–202.

[66] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make New Friends, but Keep the Old: Recommending People on Social Networking Sites," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2009, pp. 201–210.

[67] M. S. Pera and Y.-K. Ng, "A personalized recommendation system on scholarly publications," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, New York, NY, USA, 2011, pp. 2133–2136.

[68] 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 *2011*

IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011, vol. 1, pp. 96–99.

[69] M. S. Pera and Y.-K. Ng, "Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles," *J. Intell. Inf. Syst.*, vol. 42, no. 3, pp. 371–391, 2014.

[70] D. H. Lee and P. Brusilovsky, "Using Self-defined Group Activities for Improving recommendations in Collaborative Tagging Systems," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, New York, NY, USA, 2010, pp. 221–224.

[71] D. H. Lee and P. Brusilovsky, "Exploring social approach to recommend talks at research conferences," in *2012 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012, pp. 157–164.

[72] P. Avesani, P. Massa, and R. Tiella, "A Trust-enhanced Recommender System Application: Moleskiing," in *In SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, 2004, pp. 1589–1593.

[73] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proceedings of the 2007 ACM conference on Recommender systems*, New York, NY, USA, 2007, pp. 17–24.

[74] J. O'Donovan and B. Smyth, "Trust in recommender systems," in *Proceedings of the 10th international conference on Intelligent user interfaces*, New York, NY, USA, 2005, pp. 167–174.

[75] J. Golbeck, in *ISWC2005 Posters & Demostrations*, 2005, p. PID-72.

[76] 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*, New York, NY, USA, 2009, pp. 397–406.

[77] I. Esslimani, A. Brun, and A. Boyer, "From Social Networks to Behavioral Networks in Recommender Systems," in *Social Network Analysis and Mining, 2009. ASONAM '09. International Conference on Advances in*, 2009, pp. 143–148.

[78] J. A. Al-Sharawneh and M. Williams, "Credibility-aware Web-based Social Network Recommender: Follow the Leader," Jan. 2010.

[79] E. Davoodi, M. Afsharchi, and K. Kianmehr, "A social network-based approach to expert recommendation system," in *Proceedings of the 7th international conference on Hybrid Artificial Intelligent Systems - Volume Part I*, Berlin, Heidelberg, 2012, pp. 91–102.

[80] C. Chen, C. Mao, Y. Tang, G. Chen, and J. Zheng, "Personalized Recommendation Based on Implicit Social Network of Researchers," in *Pervasive Computing and the Networked World*, Q. Zu, B. Hu, and A. Elçi, Eds. Springer Berlin Heidelberg, 2012, pp. 97–107.

[81] T. Zhao, J. Hu, P. He, H. Fan, M. Lyu, and I. King, "Exploiting homophily-based implicit social network to improve recommendation performance," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 2539–2547.

[82] 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 Third ACM Conference on Recommender Systems, New York, NY, USA, 2009, pp. 53–60.

[83] D. Mican, L. Mocean, and N. Tomai, "Building a Social Recommender System by Harvesting Social Relationships and Trust Scores between Users," in *Business Information Systems Workshops*, W. Abramowicz, J. Domingue, and K. Węcel, Eds. Springer Berlin Heidelberg, 2012, pp. 1–12.

[84] P. Kazienko, K. Musial, and T. Kajdanowicz, "Multidimensional Social Network in the Social Recommender System," *IEEE Trans. Syst. Man Cybern. Part Syst. Hum.*, vol. 41, no. 4, pp. 746–759, Jul. 2011.

[85] Y. Wang and J. Vassileva, "Trust and reputation model in peer-to-peer networks," in *Third International Conference on Peer-to-Peer Computing, 2003. (P2P 2003). Proceedings*, 2003, pp. 150–157.

[86] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, "Recommender Systems Handbook," *Media*, vol. 54, no. 11, pp. 217–253, 2011.

[87] I. Esslimani, A. Brun, and A. Boyer, "Enhancing collaborative filtering by frequent usage patterns," in *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the*, 2008, pp. 180–185.

[88] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, "Recommending academic papers via users' reading purposes," in *Proceedings of the sixth ACM conference on Recommender systems*, New York, NY, USA, 2012, pp. 241–244.

[89] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th international conference on World wide web*, New York, NY, USA, 2010, pp. 421–430.

[90] Y. Zhang, J. Callan, and T. Minka, "Novelty and redundancy detection in adaptive filtering," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, 2002, pp. 81–88.

[91] K. Chandrasekaran, S. Gauch, P. Lakkaraju, and H. P. Luong, "Concept-Based Document Recommendations for CiteSeer Authors," in *Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, Berlin, Heidelberg, 2008, pp. 83–92.

[92] K. Sugiyama and M.-Y. Kan, "Scholarly paper recommendation via user's recent research interests," in *Proceedings of the 10th annual joint conference on Digital libraries*, New York, NY, USA, 2010, pp. 29–38.

[93] P. Lakkaraju, S. Gauch, and M. Speretta, "Document similarity based on concept tree distance," in *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, New York, NY, USA, 2008, pp. 127–132.

[94] K. Sugiyama and M.-Y. Kan, "Serendipitous recommendation for scholarly papers considering relations among researchers," in *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, New York, NY, USA, 2011, pp. 307–310.

[95] S. Watanabe, T. Ito, T. Ozono, and T. Shintani, “A paper recommendation mechanism for the research support system Papits,” in *International Workshop on Data Engineering Issues in E-Commerce, 2005. Proceedings*, 2005, pp. 71–80.

[96] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*. 1999.

[97] C. Hess, K. Stein, and C. Schlieder, “Trust-enhanced visibility for personalized document recommendations,” in *Proceedings of the 2006 ACM symposium on Applied computing*, New York, NY, USA, 2006, pp. 1865–1869.

[98] F. Casati, M. Marchese, K. Mirylenka, and A. Ragone, “Reviewing peer review: a quantitative analysis of peer review,” Departmental Technical Report DISI-10-014, 2010.

[99] H. Xue, J. Guo, Y. Lan, and L. Cao, “Personalized paper recommendation in online social scholar system,” in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2014, pp. 612–619.

[100] Q. Zhou, X. Chen, and C. Chen, “Authoritative Scholarly Paper Recommendation Based on Paper Communities,” in *2014 IEEE 17th International Conference on Computational Science and Engineering (CSE)*, 2014, pp. 1536–1540.

[101] J. Ha, S.-H. Kwon, and S.-W. Kim, “On Recommending Newly Published Academic Papers,” in *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, New York, NY, USA, 2015, pp. 329–330.

[102] M. Gori and A. Pucci, “Research Paper Recommender Systems: A Random-Walk Based Approach,” in *IEEE/WIC/ACM International Conference on Web Intelligence, 2006. WI 2006*, 2006, pp. 778–781.

[103] J. Ha, S.-H. Kwon, S.-W. Kim, C. Faloutsos, and S. Park, “Top-N Recommendation Through Belief Propagation,” in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2012, pp. 2343–2346.

[104] S. Pohl, F. Radlinski, and T. Joachims, “Recommending related papers based on digital library access records,” in *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, 2007, pp. 417–418.

[105] B. W. Chenxing Yang, “CARES: a ranking-oriented CADAL recommender system.,” in *Proceedings of the 2009 Joint International Conference on Digital Libraries JCDL*, Austin, USA, 2009, pp. 203–212.

[106] T. Bogers and A. van den Bosch, “Recommending scientific articles using citeulike,” in *Proceedings of the 2008 ACM conference on Recommender systems*, New York, NY, USA, 2008, pp. 287–290.

[107] D. Parra-Santander and P. Brusilovsky, “Improving Collaborative Filtering in Social Tagging Systems for the Recommendation of Scientific Articles,” in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010, vol. 1, pp. 136–142.

[108] M. Zhang, W. Wang, and X. Li, “A Paper Recommender for Scientific Literatures Based on Semantic Concept Similarity,” in *Digital Libraries: Universal and Ubiquitous Access*

to Information, G. Buchanan, M. Masoodian, and S. J. Cunningham, Eds. Springer Berlin Heidelberg, 2008, pp. 359–362.

[109] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2011, pp. 448–456.

[110] A. Vellino and D. Zeber, "A Hybrid, Multi-dimensional Recommender for Journal Articles in a Scientific Digital Library," in *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, Washington, DC, USA, 2007, pp. 111–114.

[111] D. M. Roderic, "iPhylo: Social citations: using Mendeley API to measure citation readership," *iPhylo*, Aug-2010. [Online]. Available: http://iphylo.blogspot.ca/2010/08/social-citations-using-mendeley-api-to.html. [Accessed: 04-Apr-2013].

[112] Z. Huang, W. Chung, T.-H. Ong, and H. Chen, "A graph-based recommender system for digital library," in *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, 2002, pp. 65–73.

[113] T. Y. Tang and G. McCalla, "A Multidimensional Paper Recommender: Experiments and Evaluations," *IEEE Internet Comput.*, vol. 13, no. 4, pp. 34–41, Aug. 2009.

[114] "Ex Libris the bridge to knowledge, bX Recommender Service." .

[115] J. Beel, S. Langer, M. Genzmehr, and A. Nürnberger, "Introducing Docear's research paper recommender system," in *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, 2013, pp. 459–460.

[116] D. Lee and P. Brusilovsky, "Recommending Talks at Research Conferences Using Users' Social Networks," *Int. J. Coop. Inf. Syst.*, vol. 23, no. 2, p. 1441003, May 2014.

[117] B. P. Knijnenburg, N. J. Reijmer, and M. C. Willemsen, "Each to his own: how different users call for different interaction methods in recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 141–148.

[118] Y. Hijikata, Y. Kai, and S. Nishida, "The Relation Between User Intervention and User Satisfaction for Information Recommendation," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, New York, NY, USA, 2012, pp. 2002–2007.

[119] D. Parra and P. Brusilovsky, "User-controllable personalization: A case study with SetFusion," *Int. J. Hum.-Comput. Stud.*, vol. 78, pp. 43–67, 2015.

[120] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, and A. Kobsa, "Inspectability and control in social recommenders," in *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 43–50.

[121] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "TasteWeights: A Visual Interactive Hybrid Recommender System," in *Proceedings of the Sixth ACM Conference on Recommender Systems*, New York, NY, USA, 2012, pp. 35–42.

[122] A. Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods," *User Model. User-Adapt. Interact.*, vol. 20, no. 5, pp. 383–453, Dec. 2010.

[123] D. W. Stewart and P. N. Shamdasani, *Focus groups: theory and practice*. Sage Publications, 1990.

[124] N. Tintarev, "Explanations of recommendations," in *Proceedings of the 2007 ACM conference on Recommender systems*, New York, NY, USA, 2007, pp. 203–206.

[125] J. Tang, H. Gao, X. Hu, and H. Liu, "Context-aware Review Helpfulness Rating Prediction," in *Proceedings of the 7th ACM Conference on Recommender Systems*, New York, NY, USA, 2013, pp. 1–8.

[126] B. Liu, "Informational Retrieval and Web Search," in *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*, New York: Springer:, 2007, pp. 183–236.

[127] T. Dunning, "Accurate Methods for the Statistics of Surprise and Coincidence," *Comput Linguist*, vol. 19, no. 1, pp. 61–74, Mar. 1993.

[128] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Commun ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975.

[129] C. Nascimento, A. H. F. Laender, A. S. da Silva, and M. A. Gonçalves, "A Source Independent Framework for Research Paper Recommendation," in *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, New York, NY, USA, 2011, pp. 297–306.

[130] H. Scheffé, *The Analysis of Variance*. New York: Wiley, 1959.

[131] L. Tang and H. Liu, "Relational Learning via Latent Social Dimensions," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2009, pp. 817–826.

[132] M. Granovetter, "The Strength of Weak Ties: A Network Theory Revisited," in *Sociological Theory*, 1982, pp. 105–130.

[133] K. Sugiyama and M.-Y. Kan, "'Towards Higher Relevance and Serendipity in Scholarly Paper Recommendation' by Kazunari Sugiyama and Min-Yen Kan with Martin Vesely As Coordinator," *SIGWEB Newsl*, no. Winter, p. 4:1–4:16, Feb. 2015.

[134] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, New York, NY, USA, 2010, pp. 257–260.

[135] B. Smyth and P. McClave, "Similarity vs. Diversity," in *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, London, UK, UK, 2001, pp. 347–361.

[136] M. Zhang and N. Hurley, "Avoiding Monotony: Improving the Diversity of Recommendation Lists," in *Proceedings of the 2008 ACM Conference on Recommender Systems*, New York, NY, USA, 2008, pp. 123–130.

[137] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving Recommendation Lists Through Topic Diversification," in *Proceedings of the 14th International Conference on World Wide Web*, New York, NY, USA, 2005, pp. 22–32.

Appendix A

Behavioral Ethics Certificate

UNIVERSITY OF SASKATCHEWAN

Behavioural Research Ethics Board

Certificate of Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	BEH#
Julita Vassileva	Computer Science	14-227

INSTITUTION(S) WHERE RESEARCH WILL BE CONDUCTED
University of Saskatchewan

STUDENT RESEARCHER(S)
Shaikhah Alotaibi

SPONSOR(S)
INSTITUTE OF PUBLIC ADMINISTRATION, RIYADH, SAUDI ARABIA

TITLE
Multiple Ratings of Papers Studying the Important Criteria to Include in a Paper Recommender System and How to Integrate Them

ORIGINAL REVIEW DATE	APPROVAL ON	APPROVAL OF:	EXPIRY DATE
23-Jun-2014	02-Jul-2014	Application for Behavioural Research Ethics Review Sample Email Notification Informed Consent Form Focus Group Questions	01-Jul-2015

Full Board Meeting ☐

Delegated Review ☒

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: *http://www.usask.ca/research/ethics_review/*

Beth Bilson, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1602-110 Gymnasium Place
Saskatoon SK S7N 4J8
Telephone: (306) 966-2975 Fax: (306) 966-2069

Appendix B

Consent forms

1. Consent form of the focus group study

UNIVERSITY OF SASKATCHEWAN

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

You are invited to participate in a study entitled *("multiple ratings of papers: studying the important criteria to include in paper recommender system and how to integrate them")*. Please read this form carefully, and feel free to ask the researchers any questions you might have.

Ethics Application Number: 14-227

Researchers:
Julita Vassileva, Department of Computer Science (966-2073), jiv@cs.usask.ca
Shaikhah Alotaibi, Department of Computer science, shaikhah.otaibi@usask.ca

The objective of this study to find the most important criteria to rate any research paper, and to find the how participants could combine them to evaluate the paper as a whole. In addition, the study tries to find if the users prefer to have the control over the included rating criteria in recommender system and also on the way to combine them. The data extracted from this study will be used to design a recommender system that is personalized to the user preferences.

There are no known risks in this study.

The participant will be part of a focus group that includes 6-8 people. The focus group session takes 60 to 90 minutes. The focus group will be audio recorded because each participant's opinion regarding each question is very important to the study and researchers do not like to miss anyone. Researchers will also take notes of the most important points or follow up questions that are not in the focus group plan. The data collected during the focus group does not include any personal or social information.

Personally identifying information will not be kept, and pseudonyms (alias) will be used to refer to the participants. The research data will be stored minimum of five years on a password-protected computer system and will be available only to the researchers. The researcher will undertake to safeguard the confidentiality of the discussion, but cannot guarantee that other members of the group will do so. Please respect the confidentiality of the other members of the group by not disclosing the contents of this discussion outside the group, and be aware that others may not respect your confidentiality.

Aggregate results will be used in a PhD thesis and articles published in peer reviewed conferences and scientific journals. However, any information that can be linked to a specific participant will be removed or altered.

Your participation is voluntary, and you may withdraw from the study for any reason, before the end of the data collection, without penalty of any sort. You may refuse to answer individual questions. If you withdraw from the study, any data that you have contributed will be destroyed.

If you have any questions concerning the study, please feel free to ask at any point; you are also free to contact the researchers if you have questions at a later time. This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.You may find out about the results of the study through the MADMUC website (http://madmuc.usask.ca) or by contacting the researchers.

I have read and understood the description provided above; I have been provided with an opportunity to ask questions and my questions have been answered satisfactorily. I consent to participate in the study described above and I understand that I may withdraw this consent at any time. A copy of this consent form has been given to me for my records.

I grant permission to be audio taped: Yes: ___ No: ___

(Name of Participant)	(Date)
(Signature of Participant)	(Signature of Researcher)

2. Consent form for the questionnaire

UNIVERSITY OF SASKATCHEWAN

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

You are invited to participate in a study entitled *("multi-dimensional ratings of papers: studying the differences in rating research papers between novice and experts researchers")*. Please read this form carefully, and feel free to ask the researchers any questions you might have.

Ethics Application Number: 14-227

Researchers:
Julita Vassileva, Department of Computer Science (966-2073), jiv@cs.usask.ca
Shaikhah Alotaibi, Department of Computer science, shaikhah.otaibi@usask.ca

Purpose and Procedure:

The objective of this study is to explore if there is any differences in the behavior of novice and expert researchers in rating research papers using overall ratings and multiple ratings. In addition we want to confirm if the overall rating is still important as a separate rating value or we can calculate it using the three criteria that are proposed by participants in previous study. It is also important for us to investigate the difference in the preference setting of different researchers' expertise. In this study, we will ask 5-6 experts (i.e. university professors) to recommend five papers each in a specific topic. Then the papers from all experts will be collected and send back to all of them and they will be asked to set their preferences regarding the three rating criteria: clarity, relevancy and technical soundness. In addition, experts will rate each paper using overall ratings and the three quality rating criteria. There are also few open-ended questions regarding the ratings. The same process will be done with new graduate students. All the data from experts and novice users (i.e. new graduate students) will be collected and analyzed.

Potential Benefits: the results of this study will help understanding the users' rating behaviour and some other aspects that will help in designing a paper recommender system.

Potential Risks: there are no known risks in this study.

Confidentiality: Once you sign the consent form, we will not require any personal identifiable information such as name, emails, etc. Personally identifying information will not be kept, and pseudonyms (alias) will be used to refer to the participants. The research data will be stored minimum of five years on a password-protected computer system and will be available only to the researchers.

Dissemination of Results: aggregated results will be used in a PhD thesis and articles published in peer reviewed conferences and scientific journals. However, any information that can be linked to a specific participant will be removed or altered.

Right to Withdraw: your participation is voluntary, and you may withdraw from the study for any reason, before the end of the data collection, without penalty of any sort. You may refuse to answer individual questions. If you withdraw from the study, any data that you have contributed will be destroyed.

Questions: if you have any questions concerning the study, please feel free to ask at any point; you are also free to contact the researchers if you have questions at a later time. This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.You may find out about the results of the study through the MADMUC website (http://madmuc.usask.ca) or by contacting the researchers.

Follow-Up or Debriefing: If you would like to know the results of this study, you can contact the researchers

Consent to Participate:
I have read and understood the description provided above; I have been provided with an opportunity to ask questions and my questions have been answered satisfactorily. I consent to participate in the study described above and I understand that I may withdraw this consent at any time. You can print a copy of this Consent Form for you records.

○ I consent ○ I don't consent

Appendix C

In this appendix, we show the figures for the results discussed in section 6.1.1 through 6.1.4

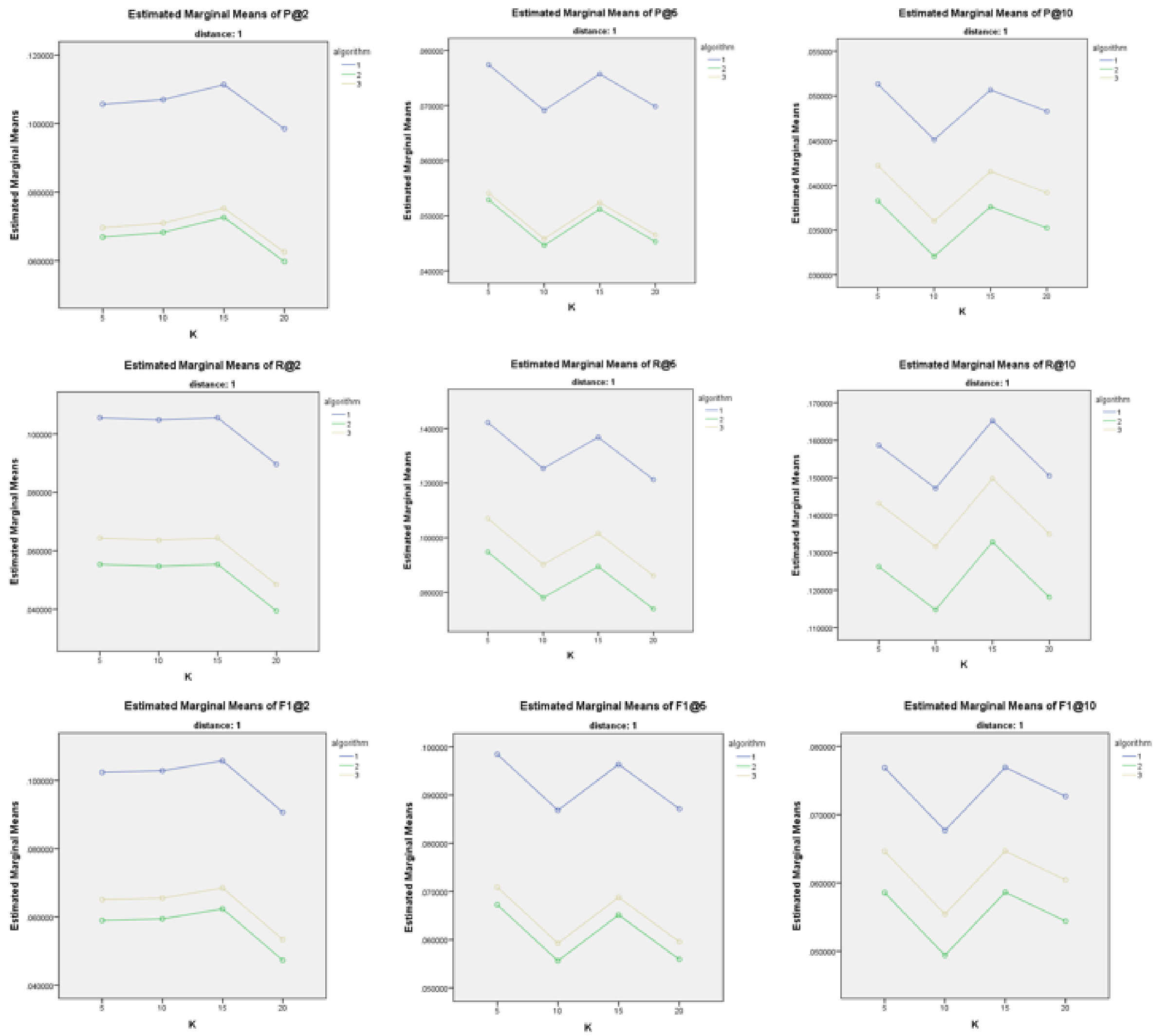


Figure C.1: Comparisons of prediction accuracy of different recommendation approaches using data from reciprocal readership ISN when one hop distance used (1-Social, 2- Combined, 3- Amplified)

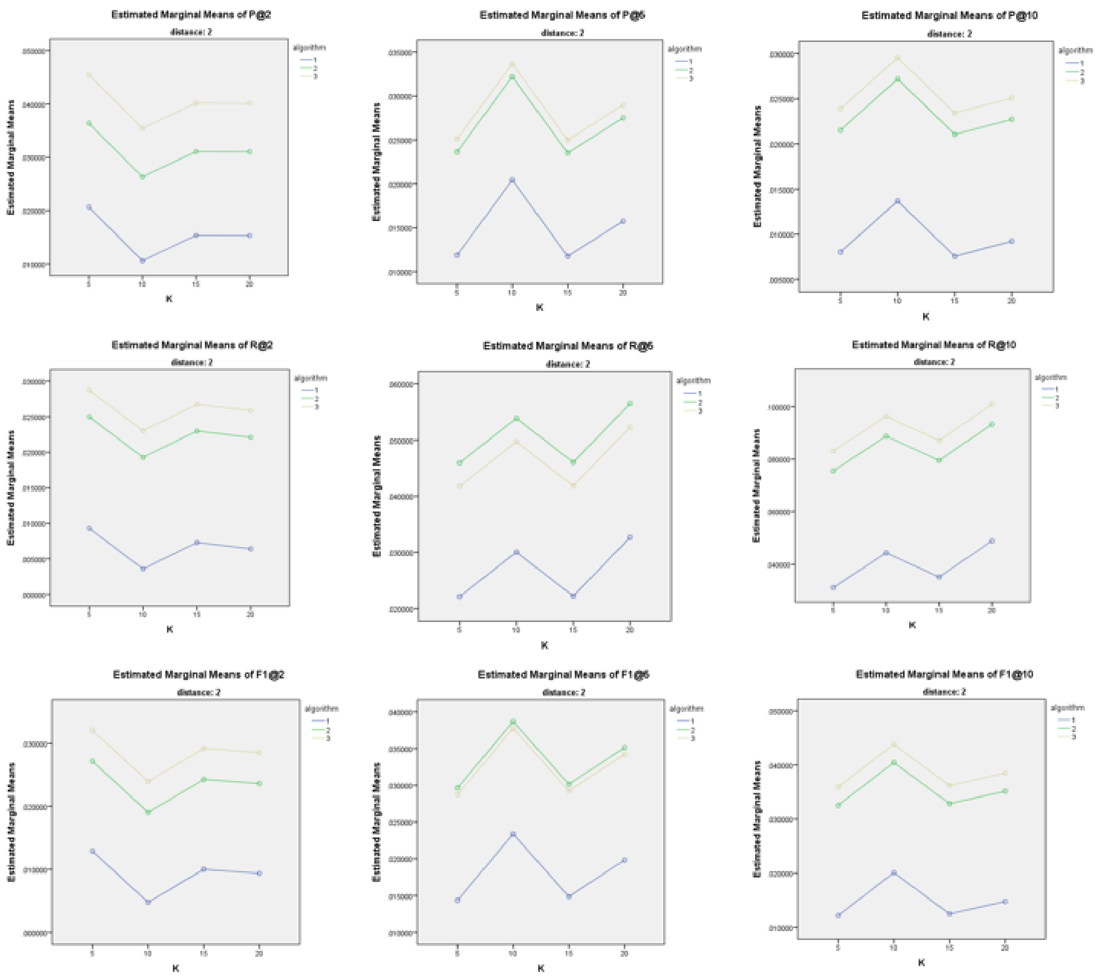


Figure C.2: Comparisons of prediction accuracy of different recommendation approaches using data from reciprocal readership ISN when two hops distance used (1-Social, 2- Combined, 3- Amplified)

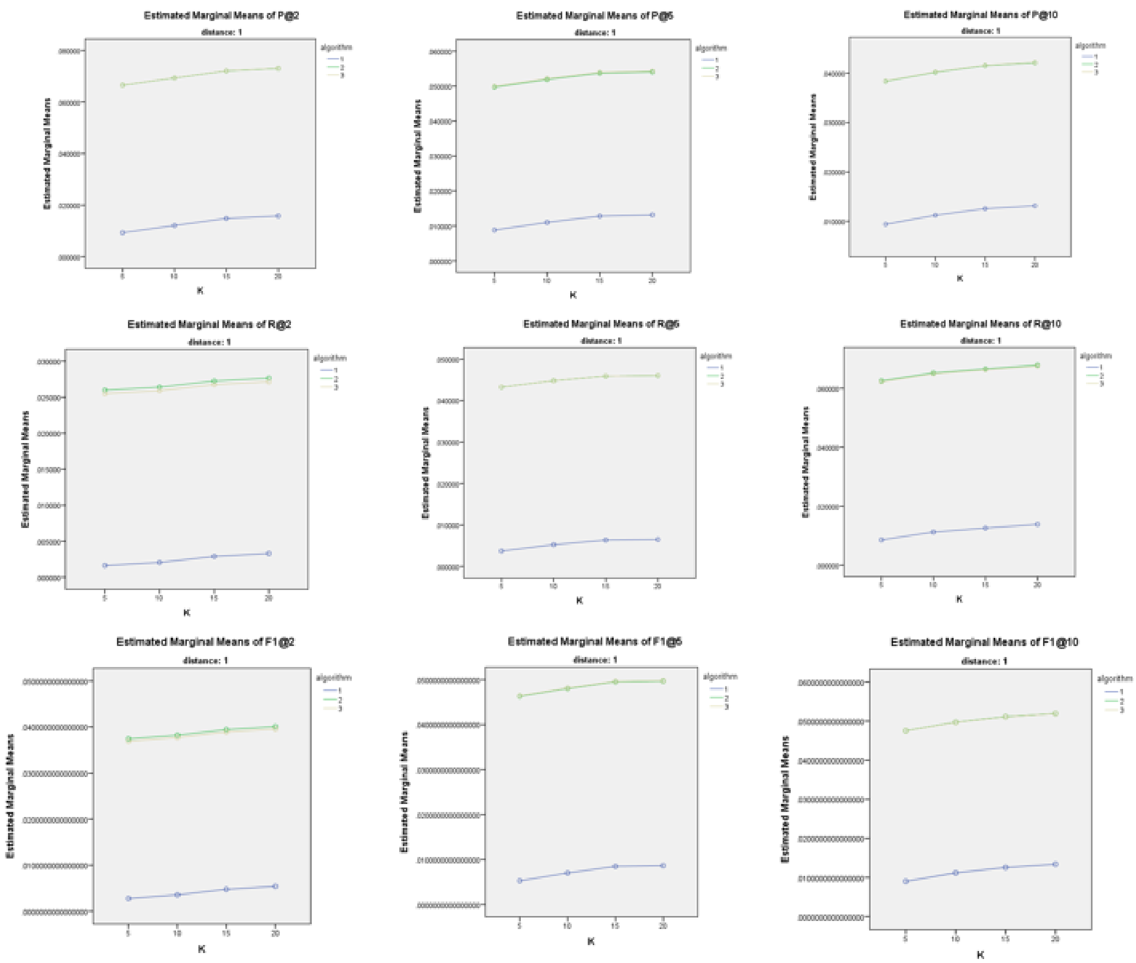


Figure C.3: Comparisons of prediction accuracy of different recommendation approaches using data from unidirectional readership ISN when one hop distance used (1-Social, 2- Combined, 3- Amplified)

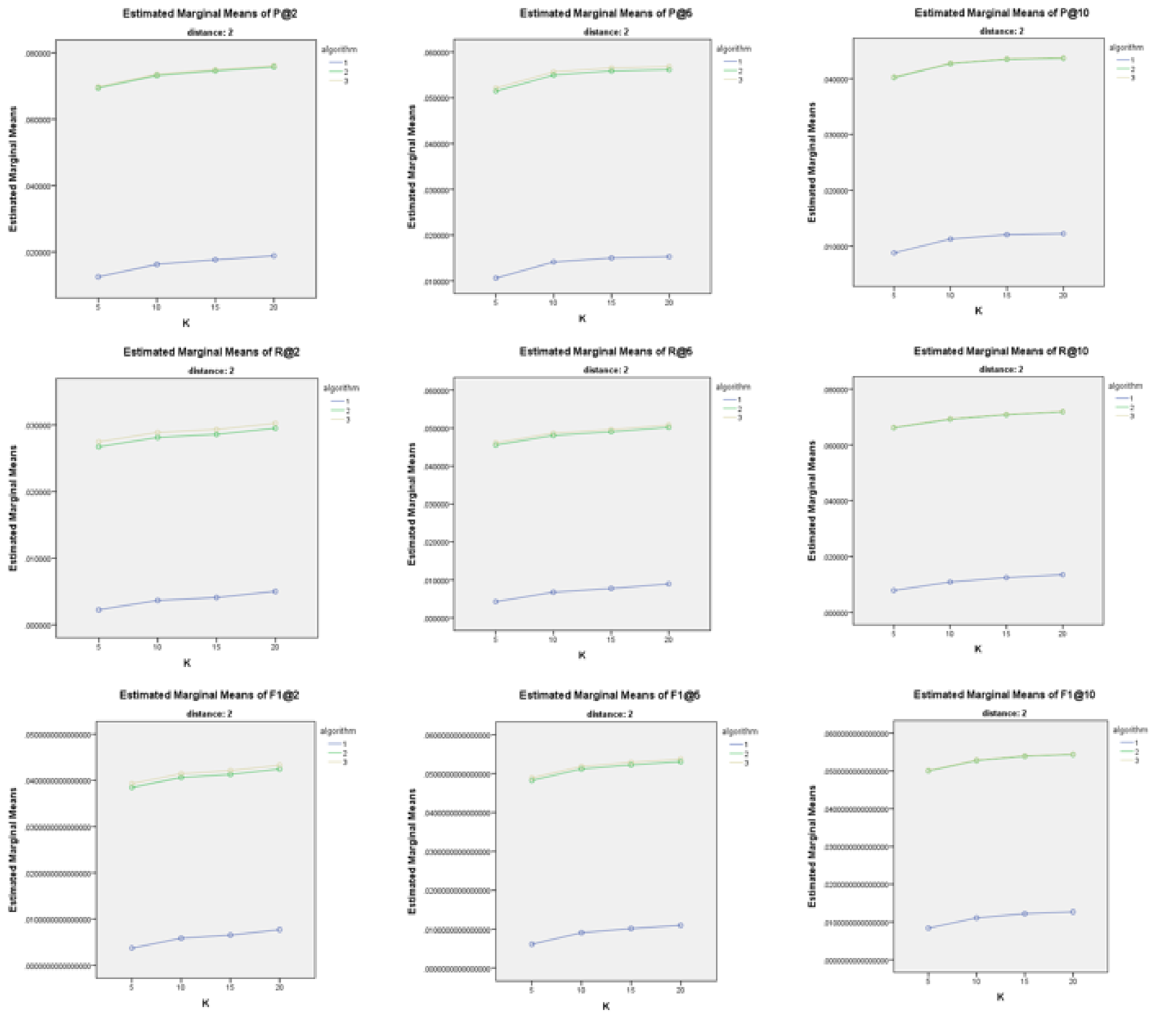


Figure C.4: Comparisons of prediction accuracy of different recommendation approaches using data from unidirectional readership ISN when two hops distance used (1-Social, 2- Combined, 3- Amplified)

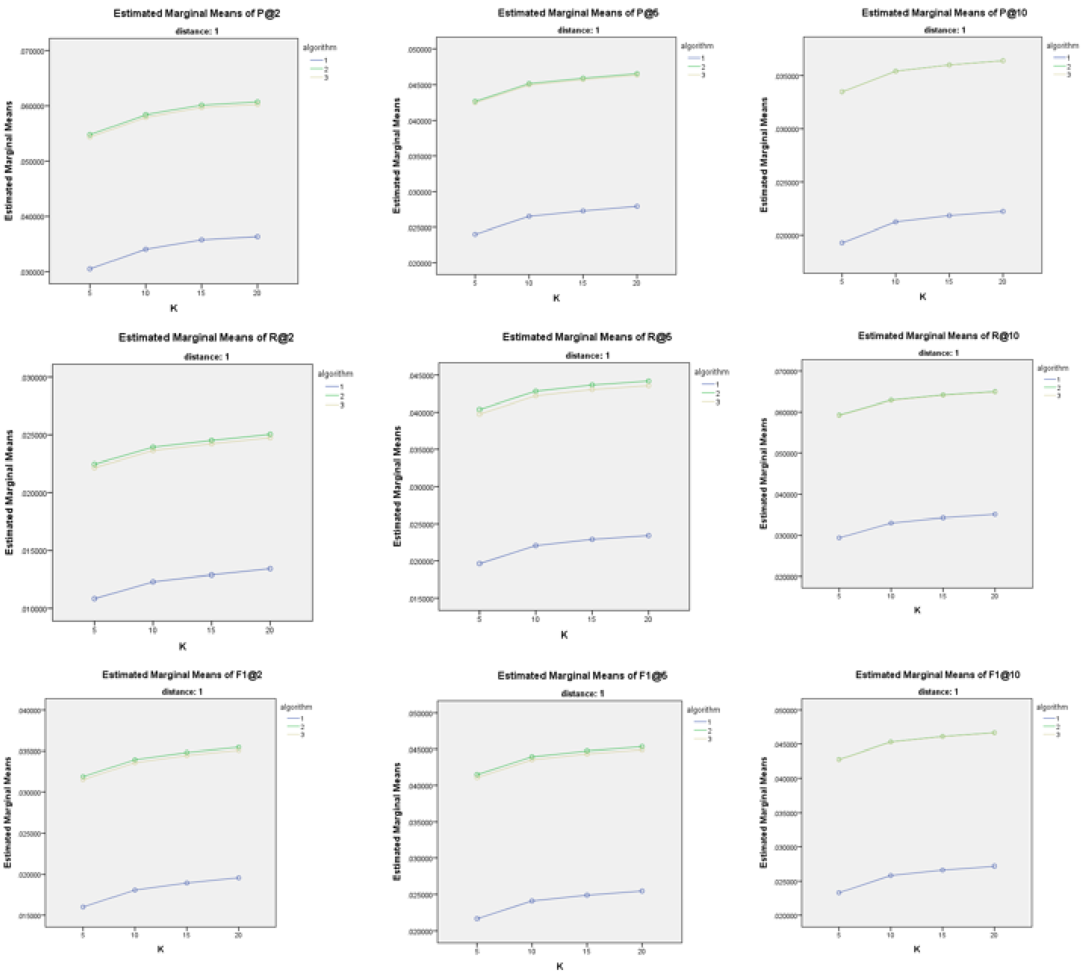


Figure C.5: Comparisons of prediction accuracy of different recommendation approaches using data from co-readership readership ISN when one hop distance used (1-Social, 2- Combined, 3- Amplified)

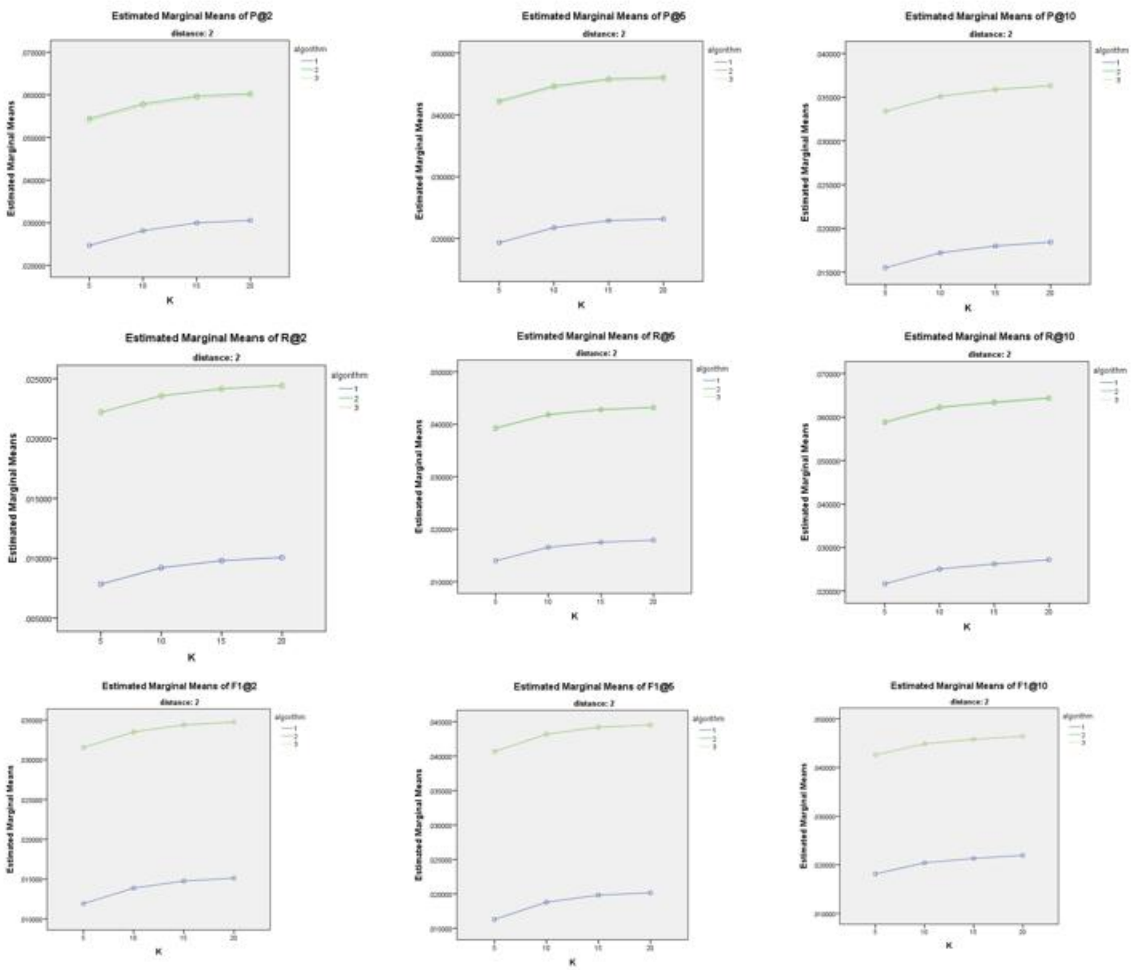


Figure C.6: Comparisons of prediction accuracy of different recommendation approaches using data from co-readership readership ISN when two hops distance used (1-Social, 2- Combined, 3- Amplified)

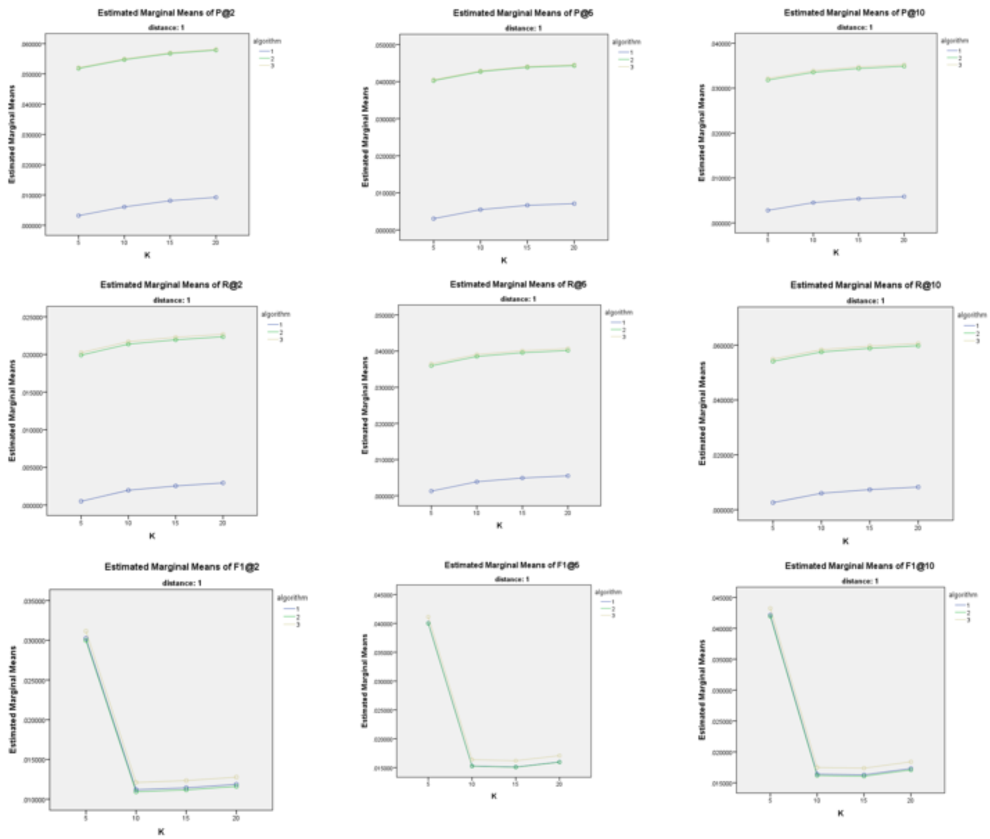


Figure C.7: Comparisons of prediction accuracy of different recommendation approaches using data from tag-based ISN when one hop distance used (1-Social, 2- Combined, 3- Amplified)

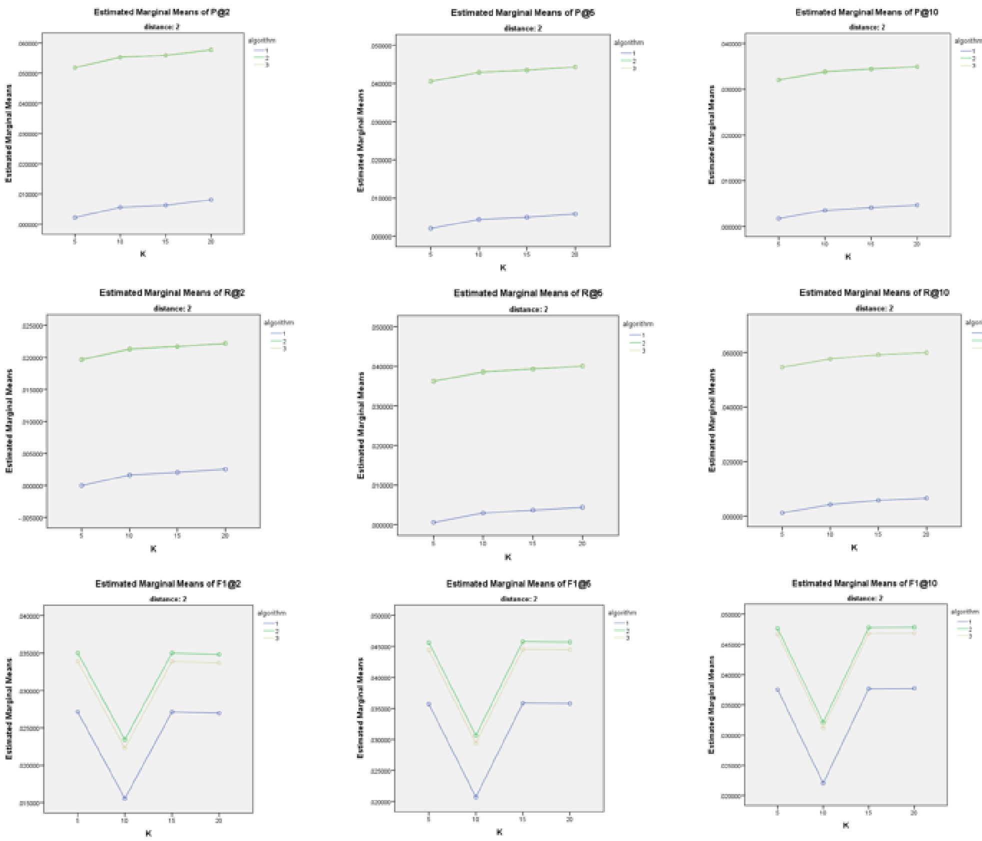


Figure C.8: Comparisons of prediction accuracy of different recommendation approaches using data from tag-based ISN when two hops distance used (1-Social, 2- Combined, 3- Amplified)