Information Research - Vol. 30 No. 3 (2025)

Cui bono? Who benefits from leveraging information behaviour of their online social connections?

Danielle Lee and Peter Brusilovsky

DOI: https://doi.org/10.47989/ir30354555

Abstract

Introduction. The underlying assumption of social, link-based information access posits that individuals engaging in online social networks derive benefits from the information shared by their online connections. This assumption is scrutinised in this study through an examination of the information-sharing behaviour of individual users with their online social connections.

Method. This research utilised a large dataset pertaining to books and movies sourced from Imhonet, integrating elements of a recommender system and a social networking service. Specifically, the study incorporated 125,657 books, 44,184 movies, associated ratings, and a social network comprising 234,789 relationships.

Analysis. The analysis of user patterns categorised the target users into two distinct categories: those exhibiting sufficiently congruent preferences with their online connections (suggesting the suitability of social link-based information access) and those lacking similar preferences with their connections (indicating potential ineffectiveness of social link-based information access).

Results. Subsequent assessments included the analysis of rating patterns and social characteristics of the target users that distinguish the two identified categories, which were examined using binary logistic regressions.

Conclusion. This study delved into the dynamics between online users' information-sharing tendencies with their online connections and the diverse user characteristics influencing these patterns.

Introduction

The plethora of social networking sites has motivated researchers from several communities to explore the use of social connections for improving information-access approaches. Within the realm of *social information access and filtering*, social links have emerged as one of the most popular sources of information, which demonstrated its value as a foundation to enhance information retrieval (Amendola et al., 2023; Brusilovsky et al., 2018), social learning through online questioning and answering (Chen et al., 2018; Yuan et al., 2020), user modelling for information personalisation (Fang et al., 2022; Li et al., 2023), and personalised recommendations (e.g., Gao et al., 2023; Ji et al., 2023; Lee and Brusilovsky, 2018; Liu et al., 2022). For instance, a range of social link-based recommendation technologies explored the use of items bookmarked, shared, tagged or rated by target users' social connections as candidate information for recommendations (Gao et al., 2023; Lee and Brusilovsky, 2018; Liu et al., 2022). Social search technologies use the search and annotation activities of socially connected users to improve the ranking and presentation of search results (Amendola et al., 2023; Brusilovsky et al., 2018; Ilkou et al., 2023). Modern media-sharing sites (e.g., YouTube, Instagram, X (formerly Twitter) and Facebook) leverage information about social connections to generate a flow of interesting items and posts for users (El-Kishky et al., 2022; Huang et al., 2020).

The key motivation behind the majority of social information access techniques is the expectation of *similar interests and tastes* between target users and their social connections. This expectation is rooted in a large volume of research on two sociality-related theories: theories of shared interests (e.g., information homophily (Chen and Deng, 2023; Furutani et al., 2023; Khanam et al., 2023)) and social influence (e.g., users' information similarities are influenced by social connections and reinforced by social cascades (Bhukya and Paul, 2023; Fang et al., 2022; Li et al., 2023)). However, the nature of social connections in online social networking and information access is considerably different from traditional social connections explored by research on homophily and social influence.

Lee and Brusilovsky (2018) highlighted in their review on social recommender systems that a significant portion of social connections in online platforms, as identified through contemporary social access methods, are *unilateral* connections, such as trusting, watching, and following, in contrast to the *bilateral* connections traditionally studied in offline social networking research. The dynamics of *unilateral* connections differ substantially from those of traditional bilateral social connections. For instance, users in systems like Epinions who *trust* another user's reviews or individuals on Twitter who *follow* a social influencer may not necessarily share similar interests and preferences with their unilateral social connections. While there may be instances where socially connected users exhibit similarities in interests, this is not always the case for many others. Consequently, the indiscriminate utilization of all social connections to enhance social information access techniques without ensuring the similarity of interests and tastes among socially connected users could yield subpar outcomes for users whose preferences diverge from those of their unilateral connections.

To cater to the diverse needs of all users, social information access methods must be able to identify which users *can benefit* from information about the behaviour (such as rating, sharing, and liking) of those they follow, as opposed to users for whom this information might introduce more confusion than clarity. The ability to differentiate these two groups of users is crucial for social information access systems to tailor their approaches, for example, by employing "switching" hybridization (Burke, 2007; LeBlanc et al., 2024), to better serve users with varying preferences, ultimately enhancing performance and user satisfaction.

This study aims to reassess the assumption of interest/taste similarity among users involved in unilateral social connections. By analysing a large dataset obtained from the social platform Imhonet, which is based on unilateral social connections, the research seeks to address two fundamental issues. First, it investigates whether individuals who established unilateral social connections (i.e., by following other users) exhibit information preferences that are sufficiently close enough to those they follow to effectively benefit from information access strategies, while utilising social connections and leverage the information behaviour of users they follow. Secondly, it explores methods to differentiate users who lack adequate similarity in preferences with their followees and therefore may *not benefit* from social link-based information access, from users who do share similar preferences with their followees and thus *can benefit* from it. The main questions being addressed in this study can be outlined as follows:

1) Do users participating in unilateral social networks have similar information preferences to the users they follow?

2) What are the key traits that distinguish users who do have similar information preferences to the social connections they follow (and thus can benefit from leveraging their information behaviour) from those who do not?

To address these questions, Imhonet users were categorised into two distinct groups: the first group comprising users with significant levels of shared interests who could leverage social information access, and the second group consisting of users without similar interests who might be better served by non-social information access strategies. Subsequently, binary logistic regression analyses were conducted to evaluate the extent to which various factors derived from user rating patterns and their social network positioning could be utilised to predict the group to which a given user is likely to belong.

Related work

Online social networks and information similarity

Table 1 presents a compilation of research studies that explore the concept of information similarity within the context of online social interactions. These studies specifically examine the alignment of information-related activities among individuals who are socially connected, aiming to demonstrate the significance of online social networks as valuable sources of information. Notably, these studies investigate the use of *unilateral relations* as information sources. Unilateral social networks have emerged as a new form of online communication in various social networking sites; users initiate relationships based on their interests in other users' information rather than personal social connections, and these connections do not require reciprocal confirmation. Consequently, the primary objective of unilateral connections is for online users to cultivate repositories of useful or intriguing information, rather than to foster personal relationships or friendships.

Related Works	Similarity Measures	Kinds of Social Networks	Domains
Anderson et al. (2012)	Co-occurrence of same user actions	Implicit online social networks and following network	Wikipedia, Stack Overflow, and Epinions
Baharifard and Motaghed (2024)	Similarity measures based on heterogenous networks with embedding methods	Citation networks and tweet-retweet relations on X (Twitter)	DBLP, X (Twitter) and IMDb
Bischoff (2012)	Similarity of music listening history	Following network	Music (Last.fm)
Hajian and White (2011)	The frequency of the same information-related activities	Following network	Online social networking site (FriendFeed)
Hong et al. (2021)	Similarity of videos	Following network	Short-video social mobile application (Douyin a.k.a., Tic Tok)
Jiang et al. (2023)	Similarity of content	Network	Microblogging Site (Twitter), especially about Covid-19 Outbreak and 2020 US Presidential Elections
Lee and Brusilovsky (2017)	Similarity of commonly bookmarked articles	Watching network	Social bookmarking system (Citeulike)
Liu et al. (2014)	Similarity of ratings on various products	Trust-based network	Product rating (Epinions.com)
Ma (2014)	Similarity of ratings on movies	Friendship and trust-based network	Movies and venues
Mi et al. (2022)	Similarity in sentiments of Tweets.	Following network, combined with 'retweets' and 'like' interactions	Microblogging sites (Sina-Weibo)
Xu et al. (2018)	Similarity of users' online networks	Following and trust-based networks	Microblogging sites (X and Sina-Weibo) and product rating (Epinions.com)
(Yin et al., 2022)	Similarity in sentiments of tweets, especially negative sentiments	Following network, derived from 'forwarding' activities	Microblogging sites (Sina-Weibo)
Zhao et al. (2020)	Similarity of ratings on various products	Trust-based network	Product ratings (Ciao-DVD and Epinions.com)
Zhu et al. (2023)	Similarity in sentiments of comments on videos and users' multidimensional interests on videos	Network built based on comment and mention relations	Short-video social mobile application (Tic Tok)

Table 1. Existing studies on information similarity of online social networks

Addressing user differences in information access technologies

It has been long recognised that, for different users, optimal personalization could be provided by compounding varying sources of information altogether (Burke, 2007; LeBlanc et al., 2024; Quadrana et al., 2018). As one of such approaches to maximise the values of different information sources in personalizing information for individual users, many researchers suggested AI-driven personalization with user control. For example, in the area of recommender systems, positive results were achieved by allowing users to choose individual peers for recommender algorithms

(Ekstrand et al., 2015) and to control the breadth of sources for personalization (Chen et al., 2022; Sun et al., 2023), as well as to adjust the strength of individual sources in the ensemble using sliders (Sciascio et al., 2018; Tsai and Brusilovsky, 2021).

While user control approaches are known as efficient, they tend to make information access interfaces more complex, creating a barrier for less experienced users. In response to this observed problem, the most recent stream of research focused on understanding individual differences between users in an attempt to take these differences into account when delivering personalization. This stream of works combines the benefit of automatic data-driven adaptation delivered by traditional hybrid approaches with the ability to better address individual users pioneered by user-controlled personalization. Two streams of recent studies demonstrate how different dimensions of user individual differences could be modelled and used to deliver better personalization.

The first stream of studies (Cai et al., 2022; Jin et al., 2019; Millecamp et al., 2019; Tkalcic and Chen, 2015) exploit users' personal traits to develop effective user interfaces or interactive experiences in personalised information access. For instance, Jin et al. (2019) adopted users' personal traits related to human cognition and perception to design user interface widgets for a music recommender system. Similarly, Cai et al. (2022) also proposed to exploit personal traits to increase the trustworthiness of conversational recommender systems. Millecamp et al. (2019) introduced explanation interfaces for music recommendations and correlated users' personal traits collected through questionnaires (e.g., trust, novelty, user intentions, satisfaction, and confidence) with the perceived effectiveness and benefits of the explanations. Birk et al. (2015) conducted a survey to characterise online game players and provide guidelines on tailoring users' gaming experiences. However, the personal characteristics explored in these studies were identified through separate questionnaires, and it was nearly impossible to identify them by analysing the users' patterns of expressing their preferences through, for instance, ratings. As a rare exception, Wu (2017) investigated how to infer users' personal traits from their implicit behaviour on a movie information system and tried to integrate the inferred traits with an existing personalised recommendation algorithm.

The second stream of studies (Dhelim et al., 2022; Dhelim et al., 2023; Fernández-Tobías et al., 2016; Guo et al., 2023; Srivastava et al., 2019) adopt users' personal traits when modelling user preferences. For instance, Srivastava et al. (2019) evaluated to what extent user classifications made by three prominent psychographic constructs could relate to *grey sheep* users, that is, those who have unusually peculiar information tastes. The above study assumed that users' characteristics are manifested in their information preferences. Several studies focused on social commerce sites (Bugshan and Attar, 2020; Ghahtarani et al., 2020; Liu et al., 2016) examined personal factors to motivate and encourage online customers to share their items of interest with online social connections. A recent survey study (Dhelim et al., 2022) elucidated how various personal traits had been used in diverse information personalization. However, these studies do not examine the determinants that shape how individual users perceive the information of online connections.

Our work attempts to continue this stream of research by understanding and modelling critical individual differences in the context of social information access. Using data about user behaviour and their positions in online social network topology, we attempt to understand which factors could discriminate users who could benefit from personalization based on social links from users who will benefit more from traditional non-social personalization based on rating behaviour. We believe that this work is important because it could lead to better hybrid social information access approaches such as switching hybrids (Burke, 2007; LeBlanc et al., 2024; Quadrana et al., 2018), which are not blindly fusing sources of information, but consider individual user differences to choose more beneficial sources.

Research methods

Dataset

One of the challenges of research on social link-based information access is the scarcity of relevant datasets. While information systems with social connections are quite popular, there are very few datasets with sufficient scale to explore the properties of unilateral social connections. In this study, we used a unique dataset from the recently discontinued system *Imhonet* that combined features of a recommender system and a social networking site. The system allowed users to review and rate a variety of products (e.g., books, movies, songs, comics, games, TV shows, and perfumes) while also establishing unilateral social connections to other users. The system has been used extensively in Russia and its neighbouring countries such as Belarus, Armenia, Georgia, Latvia, Kazakhstan, Ukraine, and Uzbekistan. In the peak of its popularity, the system had 20 million unique visits per month. A sampling of its audience taken in this period demonstrated that Imhonet engaged a broad segment of Internet users across ages and genders (refer to Figure 1).

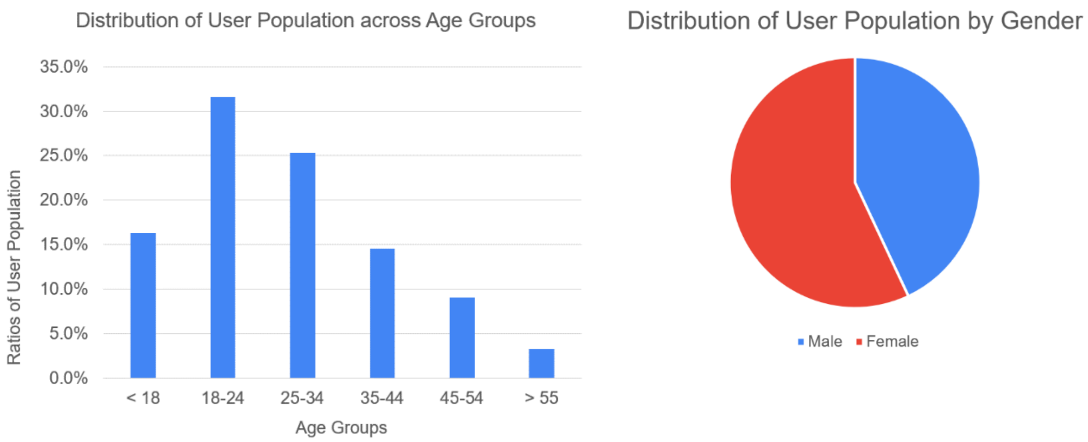

Figure 1. Distribution of user population across age groups and gender

Due to its popularity and broad representation of target audience, the system and its data had been used in a range of studies in several research areas including personalised recommendations (Arthur et al., 2022; Rodzin et al., 2020), cross-domain recommendations (Li and Tuzhilin, 2023; Sahebi and Brusilovsky, 2013), text-based machine learning (Koybagarov and Mansurova, 2016; Loukachevitch, 2021; Sharapov et al., 2019), customer segmentation for marketing (Unger et al., 2023), and audience response for media (Khitrov, 2019).

The Imhonet dataset used in this study represents data on 266,278 users and contains over 20 million book and movie ratings. The dataset includes (1) lists of users, books, and movies; (2) users' ratings of books and movies; and (3) a list of social connections. According to the rating distribution, even though there are considerably more books (125,657 books) than movies (44,184 movies), users rated movies more actively (n = 14,756,094 ratings, *Mean* = 75.7, *Standard Deviation* (hereafter, *Stdev*) = 166.8) than books (n = 8,810,045 ratings, *Mean* = 45.2, *Stdev* = 117.6).

Approximately one quarter of users in the dataset (24.27%, n = 67,760) were involved in at least one social connection establishing a total of 234,789 relationships. While these unilateral social connections were referred to in the system as *friends*, by their nature it was a typical unilateral connection, similar to *following* on X, Instagram, and TikTok. When user A finds another user B to be worth a connection for any reason, user A can connect to user B as a *friend* without B's consent. Owing to the unilateral nature of social networks, users who participated in the network can play two roles: *followers* who initiate the friend connections, and *followees* who are followed by others.

Once *followers* establish connections to other users (i.e., *followees*) on Imhonet, they can observe the activities of their *followees* in the system (i.e., rating and reviewing items) and also obtain easier access to items from these *followees*' activities. We interpret the connections of Imhonet system as *followers*' explicit expressions of interests on their *followees*' information. In other words, the information behaviour of their *followees* could serve as useful sources of information for *followers*. However, because followees' decisions to connect to their *followers* are not necessarily reciprocal, the *followees*' interests in their *followers* cannot be ascertained unless *followees* also connect to their followers. Therefore, as the target of investigation in our study, we primarily focused on 17,368 *followers* who not only participated in the unilateral social network by following at least one user, but also participated in item-related activities by rating at least one book and one movie. Throughout this study, *social connections* and *followees* were used interchangeably in order to indicate those who were followed by target users.

Measuring the information similarity between individual users and their social connections

To draw the general patterns of information sharing and more practically evaluate the usefulness of social link-based information access for each target user, in this section, we concentrated on information similarity between target users and their followees. To assess this similarity, we counted the numbers of books and movies that individual target users as well as their *followees* had co-rated and measured how similar book and movie ratings were between those users and their connections. By discerning these patterns, we were able to infer whether the target users shared comparable information preferences with their connections. This analysis is critical to enjoy social link-based information access and classify them from this perspective.

In several information access technologies such as personalised recommendations, personalised search, and social navigation, it is not easy to model users' information preferences because users do not explicitly express their information tastes. To overcome this challenge, personalised information access technologies attempt to find top peers who are the most "like-minded" with target users according to the similarity of tastes and then exploit the information possessed by the peers in further processes. Consequently, identifying top peers and measuring how similarly the peers rated the same items with target users were the core foundations of these technologies (Lee and Brusilovsky, 2018).

In accordance with this approach, we created two groups of top N peers for each of the 17,368 target users. These groups include social top N peers and non-social top N peers. *Social peers* are identified as the most similar top N connections within a target user's online social network, encompassing not only direct followees but also followees of followees (i.e., connections within a 2-hop distance). On the other hand, *non-social peers* represent the most similar top N users from the broader user population who are not closely linked and were not considered in the selection of 'social peers.' Non-social peers have social distances greater than 2 hops, including those with infinite distance (i.e., no social connection at all) from the target user.

The top 20 ranked peers were chosen as the top N rank for both social and non-social peers, aligning with previous research on personalised recommendations (Kluver, et al. (2017) explains that "past evaluations have suggested that using the 20 to 60 most similar users perform well and avoids excessive computation."). Furthermore, to explore personal and social attributes at finer level of information filtering, the top 20 peers most resembling a target user were selected instead of a larger number of peers, such as the top 60 users, which would include the 20 most similar users plus 40 less similar users.

Then, we conducted within-user comparisons by analysing two distinct groups of peers. Specifically, we calculated the average levels of similarity between target users and their social peers, as well as between target users and non-social peers, and subsequently compared these average similarity values. The aim was to assess whether one group of peers exhibited significantly more similar information preferences compared to the other group. To enhance the robustness of this analysis, we employed information similarity measures to identify the top 20 peers using four different approaches. The initial two approaches were centred on consumption similarity, which involved evaluating the shared number of books and movies rated by a target user and their social/non-social peers. The remaining two approaches focused on taste similarity, which entailed computing Pearson correlation coefficients based on users' ratings of the same books and movies (He et al., 2023; G. Jain et al., 2023).

Consequently, based on the degree of similarity between a target user's social and non-social peers, the participants were divided into two categories: *benefit* and *no-benefit*. Individuals with a higher level of information similarity with their followed users (social peers) may perceive the information shared by these connections as more valuable compared to information from non-social peers. These individuals were categorised as *benefit* users, indicating that they possess information preferences similar enough to their online connections to derive some advantage from accessing information through social links. Conversely, target users who exhibit greater information similarity with individuals who are not part of their social network may not derive significant benefits from using social connections for information access. In such cases, technologies that do not take social context into account may yield the best results. Therefore, these users were classified as *no-benefit* users.

Figure 2 presents the results, which provide several interesting observations. First, for all the ways used to compute information similarity, less than 10% of users have sufficient information similarity with their online social connections to enjoy the benefits of social link-based information access. In other words, for more than 90% of target users, the information provided by their online connections was potentially less valuable than the information provided by top-peers with low or no social associations. Second, the proportion of benefit users is higher when we calculate peers by consumption behaviour rather than by rating behaviour. Thus, socially connected users have a reasonable chance of reading the same books and watching the same movies but are less likely to agree with their connections' opinions about the consumed items.

The above results have substantiated the validity of our apprehension regarding the indiscriminate reliance on unilateral social connections for personalised information. Indeed, our analysis demonstrated that a significant number of target users did not rate the same items or expose similar preferences or tastes with their followees even though they initiated the connections and were active information consumers. Distinguishing between benefit and no-benefit groups of users poses a complex challenge. In the subsequent sections of this study, we attempt to address this issue by analysing a range of factors related to the target users' information behaviour and social network position that elucidate the reasons behind why a target user was classified into either group.

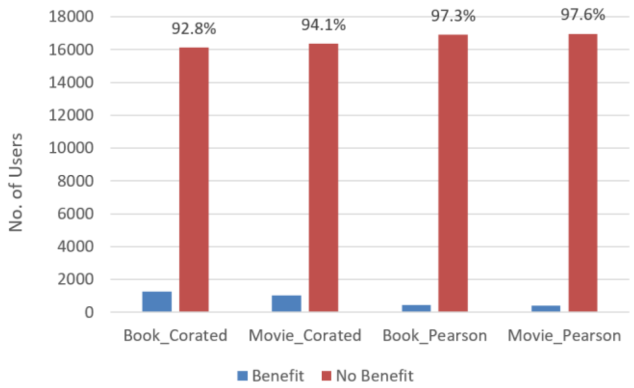

Figure 2. Number of users according to user classifications

Potential determinants to benefit/no-benefit classification

In the previous section, we measured information behaviour similarity between *Imhonet* users and their social connections and suggested an approach to classify users into *benefit* and *no-benefit* groups based on their similarity with social connections. Our next step presented in Results section is to understand how we can distinguish *benefit* and *no-benefit* users using factors that can be distilled from information available in a typical social information system, i.e., users' rating activities and social network position. In this section, we set up the context for the analysis in the Results section by introducing 11 factors (Figure 3) as potential determinants. These factors are classified into two types: (1) properties of target users' rating activities and (2) target users' topological properties in the online social network.

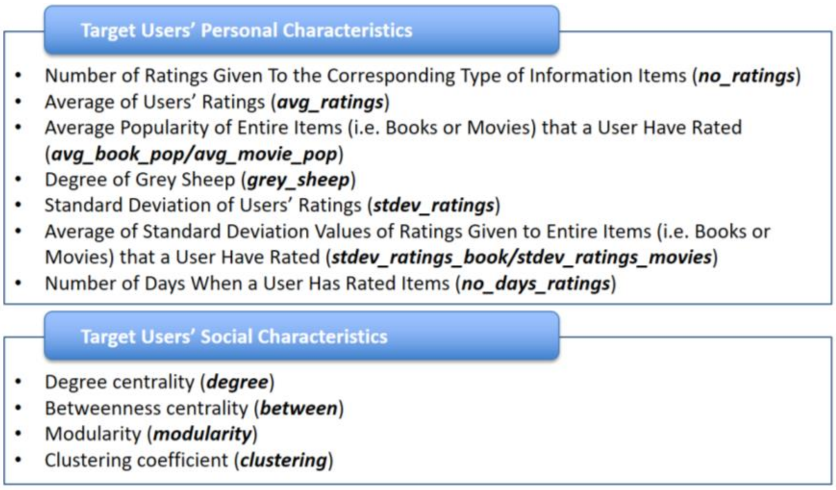

Figure 3. Factors considered in this study

The seven factors in the first type explain the properties pertinent to the rating activities of each target user as an information consumer. *The number of ratings* approximates a user's experience with the given item types (i.e., books or movies). The more items of a specific type a user has rated, the more often the user experienced items of this type (Cao et al., 2023; Moe and Trusov, 2011). *The average of a target user's ratings* represents to what extent the user is optimistic or pessimistic in evaluating items. The higher the users' average rating, the more "optimistic" they are (Jiang et al., 2019). *The average popularity of a user's items* indicates the degree to which the user aligns with prevailing trends. A user who frequently consumes and rates niche items is likely to have well-formed information preferences. By contrast, if a user has mostly consumed and rated popular items, it is likely that the user has not established a unique consumption and rating strategy for the corresponding type of items yet.

We also examined the extent to which the user's preferences could be considered as peculiar using *the "grey sheep" degree*. In contrast to the so-called *white sheep* users whose information preferences are adequately comparable with preferences of other users in the system to identify enough numbers of like-minded peers, *grey sheep* users have peculiar information preferences that make it difficult to find like-minded peers (Thorat et al., 2023). To identify *grey sheep* users, we calculated the extent of users' rating abnormality using equation (1) proposed in an earlier study (Del Prete & Capra, 2010).

$$Abnormality(u) = \frac{\sum_{i \in R_u} |r_{u,i} - \overline{r_i}|}{\|R_u\|} \tag{1}$$

where $\overline{r_i}$ is the average rating on an item i, $r_{u,i}$ is the rating of item i given by user u, R_u is the set of items rated by user u, and $\|R_u\|$ is the total number of user u's ratings.

The standard deviation of a user's ratings indicates how opinionated the user is (Mahtar et al., 2017; Ramezani et al., 2021). Highly opinionated users tend to give the lowest ratings for the items they dislike and the highest for the items they like. *The average of standard deviations of ratings given to a user's rated items* indicates how much the user tends to rate controversial items. If the average of standard deviation values of ratings given to the user's rated items is large, it suggests that the user enjoyed expressing opinions mostly on controversial items. Note that the average of standard deviation values of the user's items should not be mixed with the deviation of the user's own behaviour, such as whether the ratings assigned by the user were similar to the ratings of the other users. The latter factor is already measured by *the "grey sheep" degree* of their ratings. *The number of days a user has given ratings* reflects how regularly they used the Imhonet system since they started to use the system. The more often users visited and rated items, the more likely it was that they were exposed to the activities of their online connections. This is because users were automatically able to monitor their online connections' activities on Imhonet once they got connected.

The second set of factors aim to assess users' social characteristics by analysing their positions within the online social network. Specifically, as the utilised network is ego-centric and primarily focused on gathering valuable information, various topological properties were incorporated to capture significant aspects of users' information-related behaviour. Additionally, topologies reflecting the attributes of nodes within their immediate network surroundings were taken into account (Jia et al., 2022). Four commonly used topological properties were employed, including: 1) outdegree, 2) betweenness centrality, 3) modularity, and 4) clustering coefficients.

Outdegree refers to the number of connections followed by a target user, indicating the user's chosen information sources. *Betweenness centrality* identifies how frequently a user lies on the shortest paths between other pairs of connections. Removing users with high betweenness centrality values from the online social network could potentially elongate the paths between other connection pairs (Hajarathaiah et al., 2023). Therefore, *betweenness centrality* signifies the extent to which a user can act as an information intermediary within their online social networks (Monge and Contractor, 2003, p. 57; Tsugawa and Watabe, 2023). *Modularity* assesses the strength of a user's local community, with a high modularity value indicating strong connections within a specific community but limited connections outside of it (Peng et al., 2023; Rostami et al., 2023; Takeda and Kajikawa, 2010). The *clustering coefficient* measures the strength of a user's connections to adjacent local networks, calculated as the ratio of existing triangles of connections (i.e., cliques) to all possible triangles (L. Jain et al., 2023; Newman, 2004).

Results

This study primarily aimed to assess the factors that could affect online users' varied intentions to use their online social connections as information sources. To that end, users were operationally classified into two categories, *benefit* and *no-benefit*, based on four ways of identifying information sharing patterns, namely number of co-rated books, number of co-rated movies, the similarity of book ratings, and the similarity of movie ratings using the Pearson correlation coefficient. The classifications produced using each way were identified as four dependent variables for our analyses. The 11 potential determinants introduced in Figure 3 were identified as independent variables. Considering the dichotomous values of the dependent variables, the binary logistic regression test was chosen. It is because the logistic regression is the most typical predictive analytic methodology for dichotomous dependent variables (Brusco, 2021; Sun et al., 2022). Besides, compared with the alternative (i.e., discriminant analysis), the regression has no limitation to include both categorical and continuous independent variables as independent variables (Hassan et al., 2022; Keith, 2014, pp. 211-228). Therefore, four binary logistic regressions were executed separately for each of the four dependent variables.

Determinants affecting user classification based on the "number of co-rated books" and "number of co-rated movies"

For the first test, the classifications of target users based on the number of co-rated books and the number of co-rated movies were regressed on the 11 independent variables, and Table 2 presents the results. In the table, *B* indicates how much the dependent variables changed according to one unit change of each independent variable, and the regression equation to predict each dependent variable are made up of these *B* values as coefficients. *S.E.* in the fourth column stands for the standard errors of each independent variable. Using both S.E. values and the *Wald* values, we can estimate not only the statistical significance of each independent variables, but also the confidence interval around B values (Keith, 2014, pp. 211-228). Last Exp(*B*) represents the exponentiated version of the B. For instance, the value of 2.78 for *cluster* variable indicates that, ceteris paribus, one point increase in the *cluster* variable results in an increase in the odds of '*benefitting from friends*' *co-rated books*' by 2.78 (Brusco, 2021; Keith, 2014, pp. 211-228).

For the classification based on *co-rated books*, the regression model could make a statistically significant explanation for the benefit/no-benefit classification; $\chi^2(11) = 1363.2$, $p < 0.001$, *Negelkerke* $R^2 = .214$, ROC area = .82 (S.E. =0.008). Specifically, 21.4% variance in the classifications based on the number of co-rated books was accounted for by the independent variables. The regression model was able to correctly classify 93.2% of target users; 14.2% of target users who belonged to the *benefit* class were correctly classified as benefit, while 99.4% users belonged to *no-benefit* class were correctly classified. Thus, the model is considerably more accurate in classifying *no-benefit* users than *benefit* users.

Among the 11 potentially determining factors, six factors were statistically significant for determining users' chances of co-rating the same books with their online connections and thus potentially benefiting from information access technologies based on books rated by their online connections. To be specific, the target users' likelihood of co-rating the same books with their online connections was significantly and positively affected by *no_ratings* and *no_days_ratings*. This indicates that users who had often experienced book rating activities on the Imhonet system (i.e., users who rated many books or participated in book rating activities in more days) were more likely to pick and rate the same books as their connections. Unlike the positive directions of the two above factors, *stdev_rating_book* had significantly negative effects on target users' tendency to co-rate books with online connections. In other words, the users who rated more controversial books (higher *stdev_rating_book*) were less inclined to co-rate the same books with their online connections. By contrast, how optimistically a user rated his or her books (*avg_ratings*), how popular the rated books were (*avg_book_pop*), and how peculiar and opinionated the book ratings were (*grey_sheep* and *stdev_ratings*) compared with other users had little effect on the target user's tendency to co-rate books with online connections.

Among the three significant factors reflecting target users' social properties, the number of connections that target users followed (*outdegree*) and the structural strength of target users' local connections (*cluster*) positively affected their chance of rating the same books with their connections. In fact, among six significant factors, clustering coefficients emerged as the most notable contributors in the model (in terms of the odds ratio made by each significant variable); target users with higher clustering coefficients were 2.78 times more likely to benefit from books rated by their connections. In contrast, when a target user was a member of tightly connected communities but maintained few connections outside of the communities (higher *modularity*), the chances of benefiting from followees were lower.

		B	*SE*	*Wald*	*Exp(B)*
Classification based on No. of Co-rated Books	no_ratings	.00**	.00	100.37	1.00
	avg_ratings	.09	.05	2.70	1.09
	avg_book_pop	-.03	.06	.30	.97
	grey_sheep	-.16	.07	5.25	. 85
	stdev_ratings	.18	.07	6.05	1.20
	stdev_ratings_book	-1.3**	.29	19.25	.28
	no_days_ratings	.01**	.00	59.44	1.01
	outdegree	.01**	.00	53.51	1.01
	between	.00	.00	9.56	1.00
	modularity	-.00**	.00	43.63	1.00
	cluster	1.02**	.16	39.84	2.78
Classification based on No. of Co-rated Movies	no_ratings	.00**	.00	37.38	1.00
	avg_ratings	.12	.05	4.93	1.13
	avg_movie_pop	-.13	.06	5.11	.88
	grey_sheep	-.12	.07	2.83	.89
	stdev_ratings	.03	.09	.16	1.03
	stdev_ratings_movie	-.60*	.19	9.69	.55
	no_days_ratings	.00**	.00	48.28	1.00
	outdegree	.02**	.00	152.02	1.02
	between	.00	.00	3.04	1.00
	modularity	-.00**	.00	24.14	1.00
	cluster	1.42**	.16	75.01	4.12

* *p* <0.01, ** *p* <0.001

Table 2. Results of regression tests for the user classification based on the "number of co-rated books"

Similar results were obtained by examining factors that can explain the *benefit/no-benefit* classification of the target users based on the number of co-rated *movies*. The bottom half of Table 2 summarises the result of the regression by 11 independent variables. The regression model significantly explained the users' tendency to rate the same movies with their connections; $\chi^2(11)$ = 1249.54, p <0.001, *Negelkerke* R^2 = .214, ROC area = .83 (S.E. =0.008). Independent variables accounted for 21.4% of the variance of the dependent variable, and the classification accuracy of this model was 94.2%. Specifically, 99.6% of users whose original classification was "no-benefit" were correctly classified, and 13.0% of users whose original classification was "benefit" were correctly classified.

Resembling the results of the regression test based on books, *no_ratings* and *no_days_ratings* were significant and positive determinants; active users who rated more movies or frequently participated in rating movies were more likely to rate the same movies as their connections. Furthermore, *stdev_ratings_movie* significantly and negatively affected the chances to rate the same movies; users who tended to consume controversial movies were less likely to co-rate the same movies as their connections. The factors *avg_ratings*, *avg_movie_pop*, *grey_sheep*, and *stdev_rating* were not significant in regression focused on explaining benefit classification based on the number of co-rated movies.

The impact of factors reflecting target users' social network properties was also similar to the results of the regression test for co-rated books; the same three factors were significant. The users who followed many others (higher *outdegree*) and whose followees were more tightly connected (higher *cluster*) had a higher tendency to rate the same movies as their connections. In addition, being a member of tightly linked communities with few connections outside of the communities (higher *modularity*) lowered the odds of target users co-rating the same movies as their connections. The odds ratios showed that the clustering coefficient was also the strongest contributor among six significant factors in this co-rated movie-based regression test. The odds of co-rating the same movies with connections increased 4.12 times when the value of the clustering coefficient was increased by 1 unit.

Determinants affecting user classification based on "Pearson correlations of book ratings" and "Pearson correlations of movie ratings"

The next analysis aims to assess the effects of the 11 factors on the target users' tendency to have similar book and movie ratings with their online connections. First, the classifications of target users based on the PCC of book ratings were regressed on the 11 independent variables. Table 3 presents the results and has the same structure as Table 2.

The target users' tendency to give books similar ratings as their online connections was significantly explained by the independent variables; $\chi^2(11)$ = 156.41, p <0.001, *Negelkerke* R^2 =0.065, ROC area = .72 (S.E. =0.02). This model accounted for 6.5% variance in the classifications of Pearson correlation of book ratings. This model correctly classified 98.3% of users; however, the classification quality of the model largely varied between classes. While 100.0% of target users belonging to *no-benefit* group were correctly classified, only 3.1% of users whose original classification was *benefit* were classified correctly. As the results demonstrate, the overall percentage of correctly classified users (either *benefit* or *no-benefit*) was higher for the model based on the similarity of book tastes (98.3%) than the model based on the number of co-rated books (93.2%).

Among the 11 independent variables, four variables significantly helped in building the models to determine target users' tendency to share similar book tastes with their connections. Specifically, the more books target users rated (higher *no_ratings*), the less likely it was that their book ratings would be similar to the ratings of their online connections. Furthermore, users' positive book preferences (higher *avg_ratings*) also increased the chances for target users to rate books similarly to their online connections.

The two significant factors reflecting target users' social network properties were outdegree and clustering coefficient. When target users followed many others (higher *outdegree*), they were more likely to share similar book tastes with their connections. Furthermore, as the most contributing factor of this model, when target users' followees were more likely to connect to one another (higher *cluster*), it yielded 2.04 times increase in the target users' tendency to share similar book preferences with their connections. As in the previous model based on the number of co-rated books, *avg_book_pop*, *grey_sheep*, *stdev_ratings*, and *between* were insignificant variables of the models. Unlike the previous model, however, *stdev_ratings_book*, *no_days_ratings*, and *modularity* became insignificant as well.

		B	*S.E.*	*Wald*	*Exp(B)*
Classification based on the PCC of Book Ratings	no_ratings	-.00**	.00	11.46	1.00
	avg_ratings	.23*	.08	7.59	1.26
	avg_book_pop	.18	.08	5.36	1.19
	grey_sheep	-.01	.11	.01	.99
	stdev_ratings	.29	.12	6.39	1.34
	stdev_rating_book	-.55	.48	1.35	.58
	no_days_ratings	.01	.00	6.08	1.01
	outdegree	.00**	.00	12.14	1.00
	between	.00	.00	.21	1.00
	modularity	.00	.00	1.53	1.00
	cluster	.71*	.27	6.93	2.04
Classification based on the PCC of Movie Ratings	no_ratings	-.01**	.00	61.15	.99
	avg_ratings	.19	.09	4.69	1.21
	avg_movie_pop	.05	.07	.55	1.05
	grey_sheep	-.10	.12	.81	.90
	stdev_ratings	.25	.14	3.37	1.28
	stdev_rating_movie	.29	.33	.74	1.33
	no_days_ratings	.02**	.00	33.10	1.02
	outdegree	.00*	.00	6.66	1.00
	between	.00	.00	.51	1.00
	modularity	-.00	.00	5.59	1.00
	cluster	.96*	.28	11.96	2.61

Note: * p <0.01, ** p <0.001

Table 3. Results of Regression Tests for the user classification based on 'PCC of Book Ratings' and 'PCC of Movie Ratings'

The final test examined the factors that significantly affected the target users' likelihood of assigning movies ratings similar to their online connections; the dependent variable of this model was the classification based on the Pearson correlation coefficient of users' movie ratings, and Table 3 presents the results. The regression model was significant with 12.1% of the variance in the dependent variable explained; $\chi^2(11) = 245.16$, $p < 0.001$, *Negelkerke* R^2 =0 .121, ROC area = 0.80 (S.E. =0.01). The overall accuracy of correctly classifying target users into either *benefit* or *no-benefit* group was 98.7%; 100.0% of users whose original classifications were *no-benefit* were correctly classified, while 3.0% of users whose original classifications were *benefit* were correctly classified.

Out of seven factors related to the characteristics of users' ratings, only two factors, *no_ratings* and *no_days_ratings*, significantly affected target users' chances to have similar movie ratings as their connections. Like the results obtained for book tastes, the users who rated *more* movies (lower *no_ratings*) had a *lower* chance of exposing similar movie tastes as their connections. However, in both models, the odds ratios of this factor were almost equal to 1; the contributions of this factor in both models were marginal. The number of days a user rated movies positively affected the user's tendency to expose similar movie tastes as connections.

Finally, the two factors concerning social properties, outdegree and clustering coefficient, significantly explained the likelihood of target users to share similar movie tastes with connections. The number of target users' followees and clustering coefficient positively affected the dependent variable. In terms of the odds ratio, the strongest contributor to affect the target users' likelihood of rating movies similarly to their connections was the clustering coefficient; the increase of the factor corresponded to an increase of 2.61 in rating similarity.

Summary and Discussion

The results of our analyses of user rating patterns in a social networking system *Imhonet* revealed that users in a unilateral social network might differ considerably in terms of their opportunity to benefit from the information behaviour of their followed social connections. This observation indicated that it is important to understand how the information available in the social networking system might help in distinguishing users who *could* and who *could not* potentially benefit from information access approaches based on social connections. This understanding could help in delivering better personalization to all kinds of users by choosing the best source of information for each user and adjusting the strength of components in an ensemble of information sources. In this study, we took the first step to contribute to this understanding by attempting to determine whether a user belongs to the "benefit" or "no-benefit" class using a range of factors extracted from users' rating behaviour and their position in the social network.

Using a large dataset from a social networking system, we split the users into *benefit* and *no-benefit* categories by comparing the information similarity between socially associated peers and non-associated peers. The users in the *benefit* category had a strong tendency to behave similarly to their followees giving them a high likelihood of benefiting from social information access technologies based on the information of their followed users. The users in the *no-benefit* category tended to behave sufficiently differently from their followees and had better chances of benefiting from traditional information access approaches powered by the information about behaviour of users who are truly similar to them rather than the users they follow.

Based on the assumption that a user's tendency to resemble the behaviour of their followees could be uncovered by examining users' individual and social characteristics extracted from their rating activities and position in the social network, we performed a binary regression analysis to connect various user-related factors with users' likelihood of sharing similar information preferences with their followees. Two types of factors were evaluated: properties of target users' rating activities and their social network properties. In our tests, we employed two types of items (books and movies) and two approaches to define similar behaviours and to classify user groups. The two types of items exhibited distinct consumption and preference patterns, such as much higher user participation in movie-rating activities yielding a higher rating density of movie ratings and a wider rating coverage of books. In spite of the different rating patterns for books and movies, we were able to distil consistent ratios of user classifications and coherent sets of independent variables that could explain users' tendency to behave similarly to their connections.

First, two independent variables reflecting users' social network characteristics, *outdegree* and *clustering coefficient*, were significant with the same effect direction in the models for books and movies. The users who followed many other users and the users whose followees were more tightly connected to one another had a higher tendency to share common items and similar information preferences with their followees. Being a member of a loosely connected community with little connections to the outside of this community decreased the user's likelihood to rate common books and movies but exhibited no significant effects on sharing similar book and movie tastes. The betweenness centrality produced no significant effects on all four models.

Out of the seven factors quantifying users' rating activities, *the number of ratings* was significant for all models. While the directions of the effects made by the factor differed for both types of items and both approaches for classifying users, the odds ratios had values almost equal to one, which indicated that the factor's contributions to the models were too trivial to cause a meaningful difference. Despite its the statistical significance, *the number of days a user participated in rating activities* did not noticeably contribute to the models, either. The user tendency *to rate controversial items* negatively affected the users' tendency to rate the *same items* with connections but had no effects on rating *similarity* with connections.

The factors related to specific information preferences, i.e., how much a user's tastes were optimistic/pessimistic, unusually peculiar, and extreme, were insignificant factors in all models. Furthermore, the users' tendency to rate popular/unpopular items did not have any effects on users' information similarity with their followees. Thus, in determining whether a given user would share the same information items and similar tastes with their connections to a sufficient extent, the details of their rating activities were barely critical.

Even though the overall classification accuracies of models based on Pearson's correlation coefficient (regardless of whether book ratings or movie ratings were used) were higher, the models based on the number of co-rated items produced more accurate classifications of *benefiting* users. Furthermore, the variances in dependent variables based on the number of co-rated items explained by the regressions were two to three times higher than the variances of Pearson's correlation-based dependent variables. This result is consistent with the conclusion of the study of an online political forum for voting on various controversial political resolves conducted by Brzozowski and colleagues. This study revealed that users were influenced by their online connections in the choice of resolves (i.e., items) to vote on. However, they did not necessarily vote similarly to their connections on those resolves (Brzozowski et al., 2008).

One might suggest that some of our results are natural and hence not critical. For instance, the more books or movies target users rated, the more likely it was for them to share common items with their connections. In addition, the more users target users followed, the higher was the probability for them to share common items and expose similar preferences. However, the magnitude of odds ratios made by these variables did not overpower other significant variables. Rather, these variables were among the weakest factors explaining users' likelihood of having similar information preferences to their connections. The most notable contributors of all four regression models were the clustering coefficients that represent the strength of users' local networks, followed by the standard deviation of ratings given to users' information items that represent the degree of controversy of the entire set of user items.

Limitations and Future Work

Finally, it is important to acknowledge the limitations of our work. Most importantly, our observations and findings were made on the basis of data from just one social networking system. This limitation is typical for the research on link-based personalization as datasets of a sufficiently large scale that include users' ratings as well as their unilateral connections are extremely rare. In our work, we took advantage of the availability of a uniquely large dataset containing approximately 280,000 users, 14.6 million movie ratings, and 8.8 million book ratings. Using this dataset, we were able to demonstrate that users engaged in unilateral social networks could differ remarkably in terms of their ability to benefit from social information access technologies based on social linking information. Furthermore, the scale of this dataset enabled us to achieve significant results in our study of factors that distinguish these user groups. While we were not able to perform the same analysis with other type of items owing to the lack of datasets, the ability to repeat the analysis for two different types of items in our dataset, books and movies, and nearly similar results obtained for these two types of items provide some evidence that our work uncovered some important phenomena that could be found in social networking systems focused on other kind of items. With this, our study suggested a blueprint for future work on examining the presence of *benefit* and *no-benefit* users and discovering factors distinguishing between these categories. We believe that further work in this direction is important to improve the quality of information access technologies. As one of the future directions, we will focus on how users can be affected by their social peers' information as a part of information cascade effects (Zhong et al., 2023) by considering various topological properties of their unilateral network.

Conclusion and practical implications

This study we used a large-scale dataset sourced from Imhonet to confirm the hypothesis that users engaged in unilateral social networks differ considerably in their opportunity to benefit from the information behaviour of their connections. With this finding, the study confirmed these differences are important to consider in designing information access technologies based on social links. Moreover, the study revealed several important factors that can distinguish users who do and do not benefit from the information behaviour of their connections. We believe that these factors could be important to the developers of social information access systems to improve the quality social link-based information access algorithms either by using hybridization strategies (Burke, 2007) or by offering user informed advise about peer selection in systems which engage user in the peer selection process (Donovan et al., 2009; O'Donovan et al., 2008).

This study also offers some broader practical implications for organizations to integrate social information access technologies into their user experience or to improve their personalised information access services. Users' characteristics influence their information preferences and shape their views on information systems (Jin et al., 2019; Srivastava et al., 2019; Tkalcic and Chen, 2015), leading many researchers to create tools for identifying these traits. However, most current tools depend on human input, such as surveys (Birk et al., 2015; Cai et al., 2022; Millecamp et al., 2019). Consequently, it is uncommon to automatically assess users' traits based on their implicit online behaviour. This research emphasises the importance of not applying social information access techniques indiscriminately to all users in online social networks; instead, using one specific example, it suggests that it could be wise to start by identifying a subgroup of users who could benefit from their social connections in a specific context. It also suggests that the selection of the user subgroup can be based on users' personal traits, which can be automatically inferred from their online activities, eliminating the need for collecting personality data.

About the authors

Danielle Lee is an Associate Professor in Business School, University of Chung-Ang University, Seoul, South Korea. She received her Ph.D. from University of Pittsburgh and her research interests include social information access, information personalization and bibliometric research. She can be contacted at leedanie@cau.ac.kr.

Peter Brusilovsky is a Professor of Information Science and Intelligent Systems at the School of Computing and Information, University of Pittsburgh where he also directs Personalized Adaptive Web Systems (PAWS) lab. His primary research areas are recommender systems, intelligent user interfaces, personalized learning, and user modeling. He an editorial board member of several journals including User Modeling and User Adapted Interaction, ACM Transactions on Intelligent Systems, and International Journal of AI in Education. He can be contacted at peterb@pitt.edu.

References

Amendola, M., Passarella, A., & Perego, R. (2023). Social search: retrieving information in online social platforms - a survey. *Online Social Networks and Media*, 36, 100254. https://doi.org/10.1016/j.osnem.2023.100254

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2012). Effects of user similarity in social media. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 703-712). Assoication for Computing Machinery. https://doi.org/10.1145/2124295.2124378

Arthur, J. K., Zhou, C., Mantey, E. A., Osei-Kwakye, J., & Chen, Y. (2022). A discriminative-based geometric deep learning model for cross domain recommender systems. *Applied Sciences*, 12(10), 5202. https://www.mdpi.com/2076-3417/12/10/5202

Baharifard, F., & Motaghed, V. (2024). Similarity enhancement of heterogeneous networks by weighted incorporation of information. *Knowledge and Information Systems*, 66, 3133-3156. https://doi.org/10.1007/s10115-023-02050-x

Bhukya, R., & Paul, J. (2023). Social influence research in consumer behavior. What we learned and what we need to learn? A hybrid systematic literature review. *Journal of Business Research*, 162, 113870. https://doi.org/10.1016/j.jbusres.2023.113870

Birk, M. V., Toker, D., Mandryk, R. L., & Conati, C. (2015). Modeling motivation in a social network game using player-centric traits and personality traits. In F. Ricci, K. Bontcheva, O. Conlan, & S. Lawless (Eds.), *User Modeling, Adaptation and Personalization* (pp. 18-30). Springer. https://doi.org/10.1007/978-3-319-20267-9_2

Bischoff, K. (2012). We love rock 'n' roll: analyzing and predicting friendship links in Last.fm. In *Proceedings of the 3rd Annual ACM Web Science Conference* (pp. 47-56). Association for Computing Machinery. https://doi.org/10.1145/2380718.2380725

Brusco, M. (2021). Logistic regression via excel spreadsheets: mechanics, model selection, and relative predictor importance. *INFORMS Transactions on Education*, 23(1), 1-11. https://doi.org/10.1287/ited.2021.0263

Brusilovsky, P., Smyth, B., & Shapira, B. (2018). Social Search. In P. Brusilovsky & D. He (Eds.), *Social Information Access: Systems and Technologies* (pp. 213-276). Springer International Publishing. https://doi.org/10.1007/978-3-319-90092-6_7

Brzozowski, M. J., Hogg, T., & Szabo, G. (2008). Friends and foes: ideological social networking. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 817 – 820). Association for Computing Machinery. https://doi.org/10.1145/1357054.1357183

Bugshan, H., & Attar, R. W. (2020). Social commerce information sharing and their impact on consumers. *Technological Forecasting and Social Change*, 153, 119875. https://doi.org/10.1016/j.techfore.2019.119875

Burke, R. (2007). Hybrid web recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds), *The adaptive web: methods and strategies of web personalization* (pp. 377-408). Springer-Verlag. https://doi.org/10.1007/978-3-540-72079-9_12

Cai, W., Jin, Y., & Chen, L. (2022). Impacts of Personal Characteristics on User Trust in Conversational Recommender Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1 - 14). Association for Computing Machinery. https://doi.org/10.1145/3491102.3517471

Cao, Z., Zhu, Y., Li, G., & Qiu, L. (2023). Consequences of information feed integration on user engagement and contribution: a natural experiment in an online knowledge-sharing community. *Information Systems Research*. 35(3), 1114 – 1136. https://doi.org/10.1287/isre.2022.0043

Chen, H., & Deng, W. (2023). Leveraging heterogeneous information based on heterogeneous network and homophily theory for community recommendations. *Electronic Commerce Research*, 23(4), 2463-2483. https://doi.org/10.1007/s10660-022-09546-8

Chen, Q. Z., Schnabel, T., Nushi, B., & Amershi, S. (2022). HINT: Integration Testing for AI-based features with Humans in the Loop. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (pp. 549-565). Association for Computing Machinery. https://doi.org/10.1145/3490099.3511141

Chen, Z., Zhang, C., Zhao, Z., Yao, C., & Cai, D. (2018). Question retrieval for community-based question answering via heterogeneous social influential network. *NEUROCOMPUTING*, 285, 117-124. https://doi.org/10.1016/j.neucom.2018.01.034

Del Prete, L., & Capra, L. (2010). differS: A mobile recommender service. In 2010 *Eleventh International Conference on Mobile Data Management* (pp. 21-26). IEEE. https://10.1109/MDM.2010.22

Dhelim, S., Aung, N., Bouras, M. A., Ning, H., & Cambria, E. (2022). A survey on personality-aware recommendation systems. *Artificial Intelligence Review*, 55(3), 2409-2454. https://doi.org/10.1007/s10462-021-10063-7

Dhelim, S., Chen, L., Aung, N., Zhang, W., & Ning, H. (2023). A hybrid personality-aware recommendation system based on personality traits and types models. *Journal of Ambient Intelligence and Humanised Computing*, 14(9), 12775-12788. https://doi.org/10.1007/s12652-022-04200-5

Donovan, J. O., Gretarsson, B., Bostandjiev, S., Hollerer, T., & Smyth, B. (2009). A Visual Interface for Social Information Filtering. In 2009 *International Conference on Computational Science and Engineering* (pp. 74-81). IEEE. https://doi.org/10.1109/CSE.2009.26

Ekstrand, M. D., Kluver, D., Harper, F. M., & Konstan, J. A. (2015). Letting Users Choose Recommender algorithms: an experimental study. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 11-18). Association for Computing Machinery. https://doi.org/10.1145/2792838.2800195

El-Kishky, A., Markovich, T., Park, S., Verma, C., Kim, B., Eskander, R., Malkov, Y., Portman, F., Samaniego, S., Xiao, Y., & Haghighi, A. (2022). TwHIN: Embedding the Twitter Heterogeneous Information Network for Personalised Recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 2842 - 2850). Association for Computing Machinery. https://doi.org/10.1145/3534678.3539080

Fang, H., Li, X., & Zhang, J. (2022). Integrating social influence modeling and user modeling for trust prediction in signed networks. *Artificial Intelligence*, 302, 103628. https://doi.org/10.1016/j.artint.2021.103628

Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., & Cantador, I. (2016). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction*, 26(2), 221-255. https://doi.org/10.1007/s11257-016-9172-z

Furutani, S., Shibahara, T., Akiyama, M., & Aida, M. (2023). Analysis of homophily effects on information diffusion on social networks. *IEEE Access*, 11, 79974-79983. https://doi.org/10.1109/ACCESS.2023.3299854

Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., Chang, J., Jin, D., He, X., & Li, Y. (2023). A survey of graph neural networks for recommender systems: challenges, methods, and directions. ACM *Trans. Recomm. Syst.*, 1(1), Article 3. https://doi.org/10.1145/3568022

Ghahtarani, A., Sheikhmohammady, M., & Rostami, M. (2020). The impact of social capital and social interaction on customers' purchase intention, considering knowledge sharing in social commerce context. *Journal of Innovation & Knowledge*, 5(3), 191-199. https://doi.org/10.1016/j.jik.2019.08.004

Guo, L., Sun, L., Jiang, R., Luo, Y., & Zheng, X. (2023). Recommendation based on attributes and social relationships. *Expert Systems with Applications*, 234, 121027. https://doi.org/10.1016/j.eswa.2023.121027

Hajarathaiah, K., Enduri, M. K., Anamalamudi, S., & Sangi, A. R. (2023). Algorithms for finding influential people with mixed centrality in social networks. *Arabian Journal for Science and Engineering*, 48(8), 10417-10428. https://doi.org/10.1007/s13369-023-07619-w

Hajian, B., & White, T. (2011). Modelling influence in a social network: metrics and evaluation. In *Proceedings of IEEE third international conference on Privacy, security, risk and trust (passat) and 2011 ieee third international conference on social computing (socialcom)* (pp. 497-500). IEEE. https://doi.org/10.1109/PASSAT/SocialCom.2011.118

Hassan, W. M., Al-Dbass, A., Al-Ayadhi, L., Bhat, R. S., & El-Ansary, A. (2022). Discriminant analysis and binary logistic regression enable more accurate prediction of autism spectrum disorder than principal component analysis. *Scientific Reports*, 12(1), 3764. https://doi.org/10.1038/s41598-022-07829-6

He, X., Zhang, Y., Feng, F., Song, C., Yi, L., Ling, G., & Zhang, Y. (2023). Addressing confounding feature issue for causal recommendation. ACM *Trans. Inf. Syst.*, 41(3), Article 53. https://doi.org/10.1145/3559757

Hong, L., Yin, J., Xia, L.-L., Gong, C.-F., & Huang, Q. (2021). Improved short-video user impact assessment method based on pagerank algorithm. *Intelligent Automation \& Soft Computing*, 29(2), 437-449. https://doi.org/10.32604/iasc.2021.016259

Huang, J.-T., Sharma, A., Sun, S., Xia, L., Zhang, D., Pronin, P., Padmanabhan, J., Ottaviano, G., & Yang, L. (2020). Embedding-based retrieval in facebook search. In *Proceedings of the 26th ACM*

SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event (pp. 2553 - 2561). Association for Computing Machinery. https://doi.org/10.1145/3394486.3403305

Ilkou, E., Tolmachova, T., Fisichella, M., & Taibi, D. (2023). CollabGraph: a graph-based collaborative search summary visualization. *IEEE Transactions on Learning Technologies*, 16(3), 382-398. https://doi.org/10.1109/TLT.2023.3242174

Jain, G., Mahara, T., & Sharma, S. C. (2023). Performance evaluation of time-based recommendation system in collaborative filtering technique. *Procedia Computer Science*, 218, 1834-1844. https://doi.org/10.1016/j.procs.2023.01.161

Jain, L., Katarya, R., & Sachdeva, S. (2023). Opinion leaders for information diffusion using graph neural network in online social networks. *ACM Transactions on the Web*, 17(2), Article 13. https://doi.org/10.1145/3580516

Ji, J., Zhang, B., Yu, J., Zhang, X., Qiu, D., & Zhang, B. (2023). Relationship-aware contrastive learning for social recommendations. *Information Sciences*, 629, 778-797. https://doi.org/10.1016/j.ins.2023.02.011

Jia, M., Alboom, M. V., Goubert, L., Bracke, P., Gabrys, B., & Musial, K. (2022). Analysing egocentric networks via local structure and centrality measures: a study on chronic pain patients. In *Proceedings of 2022 International Conference on Information Networking* (pp. 152-157). IEEE. https://doi.org/10.1109/ICOIN53446.2022.9687278

Jiang, J., Ren, X., & Ferrara, E. (2023). Retweet-bert: political leaning detection using language features and information diffusion on social networks. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 459-469), Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/icwsm.v17i1.22160

Jiang, L., Cheng, Y., Yang, L., Li, J., Yan, H., & Wang, X. (2019). A trust-based collaborative filtering algorithm for E-commerce recommendation system. *Journal of Ambient Intelligence and Humanised Computing*, 10(8), 3023-3034. https://doi.org/10.1007/s12652-018-0928-7

Jin, Y., Tintarev, N., Htun, N. N., & Verbert, K. (2019). Effects of personal characteristics in control-oriented user interfaces for music recommender systems. *User Modeling and User-Adapted Interaction*, 30, 199–249. https://doi.org/10.1007/s11257-019-09247-2

Keith, T. Z. (2014). *Multiple regression and beyond. An introduction to multiple regression and structural equation modeling*. Routledge.

Kluver, D., et al. (2017). Rating-Based Collaborative Filtering: Algorithms and Evaluation. In P. Brusilovsky & D. He (Eds.), *Social Information Access* (pp. 344–390). Springer. https://doi.org/10.1007/978-3-319-90092-6_10

Khanam, K. Z., Srivastava, G., & Mago, V. (2023). The homophily principle in social network analysis: a survey. *Multimedia Tools and Applications*, 82(6), 8811-8854. https://doi.org/10.1007/s11042-021-11857-1

Khitrov, A. (2019). Assessing the realism of police series: audience responses to the russian television series Glukhar'. *Journal of Communication Inquiry*, 43(1), 6-24. https://doi.org/10.1177/0196859918774799

Koybagarov, K. C., & Mansurova, M. Y. (2016). Automatic classification of reviews based on machine learning. *Journal of Mathematics, Mechanics and Computer Science*, 91(3), 66-74.

LeBlanc, P. M., Banks, D., Fu, L., Li, M., Tang, Z., & Wu, Q. (2024). Recommender systems: a review. *Journal of the American Statistical Association*, 119(545), 773-785. https://doi.org/10.1080/01621459.2023.2279695

Lee, D., & Brusilovsky, P. (2018). Social link-based recommendations: a review. In P. Brusilovsky & D. He (Eds.), *Social Information Access* (pp. 391-440). Springer.

Lee, D. H., & Brusilovsky, P. (2017). How to measure information similarity in online social networks: a case study of Citeulike. *Information Sciences*, 418-419, 46-60. https://doi.org/10.1016/j.ins.2017.07.034

Li, N., Gao, C., Jin, D., & Liao, Q. (2023). Disentangled modeling of social homophily and influence for social recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(6), 5738-5751. https://doi.org/10.1109/TKDE.2022.3185388

Li, P., & Tuzhilin, A. (2023). Dual metric learning for effective and efficient cross-domain recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 321-334. https://doi.org/10.1109/TKDE.2021.3074395

Liu, H., Hu, Z., Mian, A., Tian, H., & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, 56, 156-166. https://doi.org/10.1016/j.knosys.2013.11.006

Liu, L., Cheung, C. M. K., & Lee, M. K. O. (2016). An empirical investigation of information sharing behavior on social commerce sites. *International Journal of Information Management*, 36(5), 686-699. https://doi.org/10.1016/j.ijinfomgt.2016.03.013

Liu, Z., Yang, L., Fan, Z., Peng, H., & Yu, P. S. (2022). Federated social recommendation with graph neural network. ACM *Trans. Intell. Syst. Technol.*, 13(4), Article 55. https://doi.org/10.1145/3501815

Loukachevitch, N. (2021). Automatic sentiment analysis of texts: the case of Russian. *The Palgrave Handbook of Digital Russia Studies* (pp. 501-516). Springer. https://doi.org/10.1007/978-3-030-42855-6_28

Ma, H. (2014). On measuring social friend interest similarities in recommender systems. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 465-474). Association for Computing Machinery. https://doi.org/10.1145/2600428.2609635

Mahtar, S., Masrom, S., Omar, N., Khairudin, N., Rahim, S., & Rizman, Z. (2017). Trust aware recommender system with distrust in different views of trusted users. *Journal of Fundamental and Applied Sciences*, 9(5S), 168-182. https://doi.org/10.4314/jfas.v9i5s.13

Mi, C., Ruan, X., & Xiao, L. (2022). Microblog sentiment analysis using user similarity and interaction-based social relations. In I. R. Management Association (Ed.), *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1887-1904). IGI Global. https://doi.org/10.4018/978-1-6684-6303-1.ch100

Millecamp, M., Htun, N. N., Conati, C., & Verbert, K. (2019). To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 397-407). Assoication for Computing Machinery. https://doi.org/10.1145/3301275.3302313

Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3), 444-456. https://doi.org/10.1509/jmkr.48.3.444

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford University Press.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200-5205. https://doi.org/10.1073/pnas.0307545100

O'Donovan, J., Smyth, B., Gretarsson, B., Bostandjiev, S., & Höllerer, T. (2008). PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1085-1088). Assoication for Computing Machinery. https://doi.org/10.1145/1357054.1357222

Peng, Y., Zhao, Y., & Hu, J. (2023). On the role of community structure in evolution of opinion formation: a new bounded confidence opinion dynamics. *Information Sciences*, 621, 672-690. https://doi.org/10.1016/j.ins.2022.11.101

Quadrana, M., Cremonesi, P., & Jannach, D. (2018). Sequence-aware recommender systems. ACM *Computing Surveys*, 51(4), 1-36. https://doi.org/10.1145/3190616

Ramezani, M., Akhlaghian Tab, F., Abdollahpouri, A., & Abdulla Mohammad, M. (2021). A new generalised collaborative filtering approach on sparse data by extracting high confidence relations between users. *Information Sciences*, 570, 323-341. https://doi.org/10.1016/j.ins.2021.04.025

Rodzin, S., Rodzina, O., & Rodzina, L. (2020). Bio-inspired collaborative and content filtering method for online recommendation assistant systems. In *Proceedings of the 9th Computer Science On-line Conference* (pp. 110-119). Springer. https://doi.org/10.1007/978-3-030-51971-1_9

Rostami, M., Oussalah, M., Berahmand, K., & Farrahi, V. (2023). Community detection algorithms in healthcare applications: a systematic review. *IEEE Access*, 11, 30247-30272. https://doi.org/10.1109/ACCESS.2023.3260652

Sahebi, S., & Brusilovsky, P. (2013). Cross-domain collaborative recommendation in a cold-start context: the impact of user profile size on the quality of recommendation. In *Proceedings of User Modeling, Adaptation, and Personalization* (pp. 289-295). Springer. https://doi.org/10.1007/978-3-642-38844-6_25

Sciascio, C. d., Brusilovsky, P., & Veas, E. (2018). A study on user-controllable social exploratory search. In *Proceedings of 23rd International Conference on Intelligent User Interfaces* (pp. 353-364), Assoication for Computing Machinery. https://doi.org/10.1145/3172944.3172986

Sharapov, R. V., Varlamov, A. D., & Sharapova, E. V. (2019). Method for sentiment text analysis based on statistical and semantic properties of words. In *Proceedings of 2019 International Russian Automation Conference* (pp. 1-6). IEEE. https://doi.org/10.1109/RUSAUTOCON.2019.8867775

Srivastava, A., Bala, P. K., & Kumar, B. (2019). New perspectives on gray sheep behavior in E-commerce recommendations. *Journal of Retailing and Consumer Services*, 53, 1-11. https://doi.org/10.1016/j.jretconser.2019.02.018

Sun, R., Akella, A., Kong, R., Zhou, M., & Konstan, J. A. (2023). Interactive content diversity and user exploration in online movie recommenders: a field experiment. *International Journal of Human–Computer Interaction*, 40(22), 1-15. https://doi.org/10.1080/10447318.2023.2262796

Sun, R., An, L., Li, G., & Yu, C. (2022). Predicting social media rumours in the context of public health emergencies. *Journal of Information Science*, 2022, 1-16. https://doi.org/10.1177/01655515221137879

Takeda, Y., & Kajikawa, Y. (2010). Tracking modularity in citation networks. *Scientometrics*, 83(3), 783-792. https://doi.org/10.1007/s11192-010-0158-z

Thorat, S. A., Ashwini, G., & Seema, M. (2023). Survey on collaborative and content-based recommendation systems. In *Proceedings of 2023 5th International Conference on Smart Systems and Inventive Technology* (pp. 1541-1548). IEEE. https://doi.org/10.1109/ICSSIT55814.2023.10061072

Tkalcic, M., & Chen, L. (2015). Personality and recommender systems. In *Recommender systems handbook* (pp. 715-739). Springer.

Tsai, C.-H., & Brusilovsky, P. (2021). The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction*, 31(3), 591-627. https://doi.org/10.1007/s11257-020-09281-5

Tsugawa, S., & Watabe, K. (2023). Identifying influential brokers on social media from social network structure. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 842-853). Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/icwsm.v17i1.22193

Unger, M., Li, P., Sen, S., & Tuzhilin, A. (2023). Don't Need All Eggs in One Basket: Reconstructing Composite Embeddings of Customers from Individual-Domain Embeddings. ACM *Transactions on Management Information Systems*, 14(2), Article 20. https://doi.org/10.1145/3578710

Wu, W. (2017). Implicit acquisition of user personality for augmenting recommender systems. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion* (pp. 201-204). Assoication for Computing Machinery. https://doi.org/10.1145/3030024.3038287

Xu, K., Zheng, X., Cai, Y., Min, H., Gao, Z., Zhu, B., Xie, H., & Wong, T.-L. (2018). Improving user recommendation by extracting social topics and interest topics of users in uni-directional social networks. *Knowledge-Based Systems*, 140, 120-133. https://doi.org/10.1016/j.knosys.2017.10.031

Yin, F., Xia, X., Pan, Y., She, Y., Feng, X., & Wu, J. (2022). Sentiment mutation and negative emotion contagion dynamics in social media: a case study on the Chinese Sina microblog. *Information Sciences*, 594, 118-135. https://doi.org/10.1016/j.ins.2022.02.029

Yuan, S., Zhang, Y., Tang, J., Hall, W., & Cabotà, J. B. (2020). Expert finding in community question answering: a review. *Artificial Intelligence Review*, 53(2), 843-874. https://doi.org/10.1007/s10462-018-09680-6

Zhao, J., Wang, W., Zhang, Z., Sun, Q., Huo, H., Qu, L., & Zheng, S. (2020). TrustTF: a tensor factorization model using user trust and implicit feedback for context-aware recommender systems. *Knowledge-Based Systems*, 209, 106434. https://doi.org/10.1016/j.knosys.2020.106434

Zhong, C., Xiong, F., Pan, S., Wang, L., & Xiong, X. (2023). Hierarchical attention neural network for information cascade prediction. *Information Sciences*, 622, 1109-1127. https://doi.org/10.1016/j.ins.2022.11.163

Zhu, H., Wei, H., & Wei, J. (2023). Understanding users' information dissemination behaviors on Douyin, a short video mobile application in China. *Multimedia Tools and Applications*, 83, 58225–58243. https://doi.org/10.1007/s11042-023-17831-3