Contents lists available at ScienceDirect

Journal of Informetrics

Recommendation of scholarly venues based on dynamic user interests

Hamed Alhoori [a,b,*], Richard Furuta [c]

[a] *Northern Illinois University, DeKalb, IL, USA*
[b] *Argonne National Laboratory, IL, USA*
[c] *Texas A&M University, College Station, TX, USA*

ARTICLE INFO

ABSTRACT

The ever-growing number of venues publishing academic work makes it difficult for researchers to identify venues that publish data and research most in line with their scholarly interests. A solution is needed, therefore, whereby researchers can identify information dissemination pathways in order to both access and contribute to an existing body of knowledge. In this study, we present a system to recommend scholarly venues rated in terms of relevance to a given researcher's current scholarly pursuits and interests. We collected our data from an academic social network and modeled researchers' scholarly reading behavior in order to propose a new and adaptive implicit rating technique for venues. We present a way to recommend relevant, specialized scholarly venues using these implicit ratings that can provide quick results, even for new researchers without a publication history and for emerging scholarly venues that do not yet have an impact factor. We performed a large-scale experiment with real data to evaluate the current scholarly recommendation system and showed that our proposed system achieves better results than the baseline. The results provide important up-to-the-minute signals that compared with post-publication usage-based metrics represent a closer reflection of a researcher's interests.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In addition to the variety of challenges researchers face from the rising number of scholarly events and venues, the important task of identifying relevant publication opportunities is further complicated due to the expansion and overlap of what were previously discrete specializations. More and more collaboration is taking place between disciplines in the research landscape, which is leading to decreased compartmentalization overall. Increasingly complex academic sub-disciplines and emerging interdisciplinary research areas, though certainly a net gain for the community as a whole, compound this problem. In such a sophisticated research environment, researchers are finding it challenging to remain up to date on new findings, even within their own disciplines (Kuruppu & Gruber, 2006; Murphy, 2003). Furthermore, "context-drift" in scholarly communities is becoming more prevalent as researchers expand, evolve, or adapt their interests in rapidly changing subject areas.

Generally, researchers become aware of scholarly venues related to their research interests by word of mouth from lab members, departmental colleagues, and members of other scholarly communities; through online searches for scholarly

---

* Corresponding author at: Northern Illinois University, DeKalb, IL, USA.
E-mail address: alhoori@niu.edu (H. Alhoori).

material; and from rankings of venues and publishers' reputations (Buchanan, Cunningham, Blandford, Rimmer, & Warwick, 2005; Chu & Law, 2007). In the past, these approaches have yielded satisfactory results, as there were relatively few venues related to any given field. However, in today's multifaceted, diverse, and interdisciplinary scholarly environment, researchers can become acquainted with newly available and relevant specialized venues only by spending considerable time and effort explicitly searching for venues that align with their research interests.

It is also essential for funding agencies to become aware of new avenues of research across fields in order to determine future allocations. Further, new interdisciplinary research areas lead to greater challenges for research institutes as they strive to understand dynamic information needs and information-seeking behaviors. Information specialists need prompt and seamless measurements of researchers' readings in order to make decisions on venue subscriptions, instead of relying blindly on the venue's impact factor or on users' explicit requests. For example, Springer provides its users with a form for recommending journals to librarians (Springer, 2015), but this feedback represents only the interests of the individuals who submit recommendations, rather than providing a picture of the entire constituency's needs.

Many rankings of scholarly venues have been created and used to help researchers become more aware of specific scholarly communities. However, knowing that very prestigious journals, such as *Science* and *Nature,* are considered top venues for multidisciplinary fields does not help researchers seeking more specialized venues and communities. Moreover, traditional citation analysis cannot provide quick, adaptive results, especially for new scholarly venues that do not yet have an impact factor.

A number of online services provide collections of venues in an attempt to alleviate some of these problems. For example, the HCI Bibliography (Perlman, 1991) is a specialized bibliographic database on Human-Computer Interaction.[1] AllConferences and Lanyrd are global conference and event directories.[2] ConferenceAlerts,[3] EventSeer,[4] and WikiCFP provide notifications of upcoming academic events based on keywords.[5] ConfSearch (Kuhn & Wattenhofer, 2008) enables researchers to search for computer science conferences using keywords, related conferences, and authors. ConfAssist (Singh, Chakraborty, Mukherjee, & Goyal, 2016) classifies conferences as top-tier or not.

However, in this era of big data, retrieving relevant results by manually searching and browsing online is no longer the only approach to discover new information, not is it generally the most efficient approach. Studies have been conducted in an effort to offer techniques capable of accelerating scholarly discovery, such as summarization, visualization (Gove, Dunne, Shneiderman, Klavans, & Dorr, 2011), and collaborative information synthesis (Blake & Pratt, 2006). Recommender systems have been introduced to filter the overwhelming amount of data by using various data analysis techniques to alleviate information overload (Shenk, 1997; Speier, Valacich, & Vessey, 1999). Recommender systems are already entrenched in the digital landscape, as they provide millions of online users with continually updated suggestions for news, books, restaurants, tourism, movies, and television programs.

With the proliferation of publications, researchers are utilizing academic social networks and reference management systems in order to find, store, and manage references (Farooq, Song, Carroll, & Giles, 2007). Social and online reference management systems enable users to bookmark references to research content, as well as tag, review, and rate research content within their profiles. Scholarly tools such as these play an essential role in the organization of personal article collections and the generation of bibliographies across the research landscape today. Scholarly communities are sharing these digital reference libraries, and this open sharing encourages the formation of new research groups. Such online personal collections or repositories also accurately reflect researchers' current and past reading, and indicate changes in their interests over time, making these datasets prime targets for recommendation analytics.

In previous work (Alhoori, 2016; Alhoori & Furuta, 2011), we found that several of the participating researchers expressed a notable desire to be aware of new and well-established scholarly venues and events related to their shifting research interests. In this paper, we build a personal measure for evaluating venues based on user-centric altmetrics and analysis of readings, rather than relying on conventional citation-based metrics. Then, we augment the researchers' awareness and recommend semantically related scholarly venues based on their interests. In creating this measure, we draw on data from CiteULike,[6] a well-known social reference management system.

This paper is structured as follows: In Section 2, we discuss related work. In Section 3, we describe an approach for measuring an implicit rating for scholarly venues by monitoring researchers' behavior. In Section 4, we explain the data collection and the experiments. In section 5, we present and discuss the results.

2. Related work

Recommender systems streamline and augment a person's decision-making process, especially when inadequate information is available with which to make an informed decision (Resnick & Varian, 1997). One well-known recommender

---

[1] http://www.allconferences.com/
[2] http://lanyrd.com/
[3] http://www.conferencealerts.com
[4] http://eventseer.net/
[5] http://www.wikicfp.com/
[6] http://www.citeulike.org/

technique is collaborative filtering (CF) (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994; Schafer, Frankowski, Herlocker, & Sen, 2007). User-based collaborative filtering stems from the idea that users whose respective ratings show a high level of agreement and/or who have a similar history of behaviors are likely to continue to show agreement in these regards. This algorithm searches for users who share similar patterns to those of a current user and uses their ratings to predict unidentified preferences for the current user. Item-based collaborative filtering uses similarities between item ratings to predict users' preferences instead of using similarities between users' ratings (Sarwar, Karypis, Konstan, & Reidl, 2001).

Other recommender systems use a matrix factorization approach based on the stochastic gradient descent (SGD) (Bottou & Bousquet, 2008),[7] singular value decomposition (SVD) (Sarwar, Karypis, Konstan, & Riedl, 2000), or SVD++ (Koren, Bell, & Volinsky, 2009). SGD is an iterative learning algorithm for minimizing the error between actual and predicted ratings. SVD reduces the dataset by eliminating insignificant users or items. SVD++ constitutes an improvement on SGD in which it not only considers ratings but also considers who has rated what (e.g., rating an item is an indication of preference).

Recommender systems have been used to recommend movies (Wu & Niu, 2015), research papers (Beel, Gipp, Langer, & Breitinger, 2015), collaborators (Yan & Guns, 2014), experts (Protasiewicz et al., 2016), reviewers (Basu, Cohen, Hirsh, & Nevill-Manning, 2001), citations (Caragea, Silvescu, Mitra, & Giles, 2013), and tags (Song, Zhang, & Giles, 2011).

The need to connect authors and readers goes back to at least 1974 when Kochen and Tagliacozzo (1974) proposed a service to suggest journals for authors' manuscripts using a mathematical model that took into consideration relevance, acceptance rate, circulation, prestige and publication lag. However, until just a few years ago very little progress had been made in this area. Since then, due in large part to the increasing information overload researchers face when searching for new venues, there has been a resurgence in research and development surrounding the recommendation of scholarly events (Huynh & Hoang, 2012).

Klamma et al. (2009) recommended academic events based on researchers' event participation history, whereas (Luong, Huynh, Gauch, Do, & Hoang, 2012; Luong, Huynh, Gauch, & Hoang, 2012) used co-authors' publication history to recommend venues. Boukhris and Ayachi (2014) proposed a hybrid recommender for upcoming conferences related to computer science based on venues from co-authors, co-citers, and co-affiliated researchers. Pham et al. (2011) clustered users on social networks and used the number of papers a researcher had published in a venue to derive the researcher's rating for that venue. eTBLAST (Errami, Wren, Hicks, & Garner, 2007) and the Journal Article Name Estimator (Jane) (Schuemie & Kors, 2008) recommend biomedical journals based on an assessment of abstract similarity. Silva et al. (2015) considered the quality and relevance of manuscripts in order to recommend journals. They also analyzed the authors' social networks and identified journals in which similar researchers had published. Other venue recommendation approaches have based ratings on the topic and writing style of a paper (Yang & Davison, 2012), the title and abstract of a paper (Medvet, Bartoli, & Piccinin, 2014), an analysis of PubMed log data (Lu, Xie, & Wilbur, 2009), and personal bibliographies and citations (Küçüktunç, Saule, Kaya, & Çatalyürek, 2012; Kucuktunc, Saule, Kaya, & Çatalyürek, 2013).

Recently, some online services have started to provide support for locating relevant journals using title, keyword, and abstract matching. These services include Elsevier Journal Finder (Kang, Doornenbal, & Schijvenaars, 2015),[8] Springer Journal Selector,[9] EndNote Manuscript Matcher,[10] Jane,[11] and Edanz Journal Selector.[12]

In addition, more research has been carried out in recent years on recommending events in general. For example, Xia et al. (2013) presented a socially aware recommendation system for conference sessions, and Quercia et al. (2010) used mobile phone location data to recommend social events. Minkov et al. (2010) proposed an approach to recommending future events, whereas Khrouf and Troncy (2013) used hybrid event recommendations with linked data.

Most research on scholarly venue recommendation to date has used citation analysis and the publication or participation history of researchers to build recommendations. Unfortunately, this model cannot be widely generalized, as it would not be useful for new researchers or graduate students who lack an established record of scholarly activity. Furthermore, using only the venues in which a researcher has previously published work undermines the recommendation process, as a researcher might be interested in new research areas in which she or he has not yet published any articles. This research study explores pathways with the purpose of drawing on a researcher's current personal article collections and readings to build tailored venue recommendations.

3. Personal venue rating (PVR)

Venues can prove difficult to analyze for the purpose of building recommendations, as explicit metadata or user ratings are scarce and hard to come by. Research articles, however, have a variety of associated metadata fields. References in a researcher's library can thus provide easily accessible, indirect information pertaining to a researcher's interests. We used references and the years in which they were added to a researcher's library in the measurement of personal venue rating.
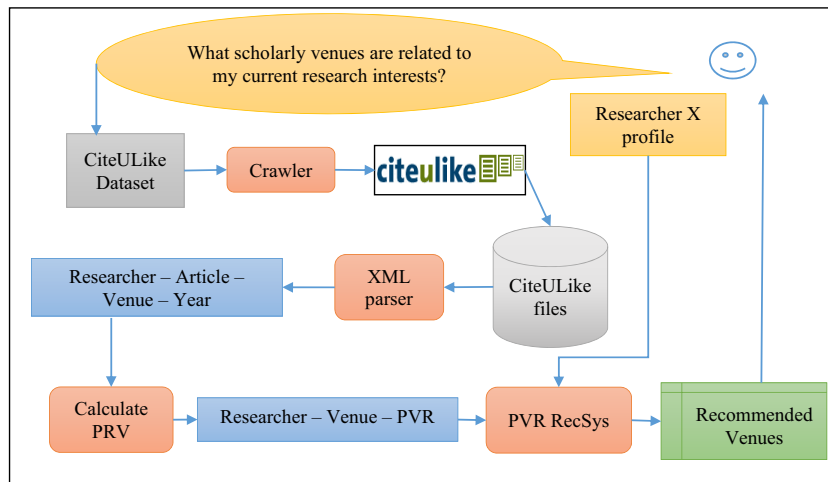
Fig. 1. The PVR recommendation system architecture.

PVR takes into consideration how a researcher's interest in a given venue has changed over time. In Eq. (1), we define *PVR* as a weighted sum for researcher *u* and venue *v*, and $r_{u,v}$ we refer to it as

$$r_{u,v} = \sum_{i=y}^{1} w\log(v_{u,i} + 1) \tag{1}$$

$v_{u,i}$ denotes the number of references in a researcher's library *u,* from a specific venue v, which the researcher added during a certain year of the total number of years y during which the researcher followed venue v. The weight w increases the importance of newly added references and is equal to i. PVR favors researchers who have followed a venue for several years over researchers who have added numerous references from a venue over fewer years. The log minimizes the effect of adding numerous references and helps to reduce the impact of potential shilling attempts (Lam & Riedl, 2004). The addition of one allows for the case of one reference to be added to a library in a year. We used the year that a reference was added to the researcher's library, as it is more relevant to the researcher than the year the article was published.

4. Data and experiments

4.1. Metrics

We conducted an offline experiment using the CiteULike dataset,[13] which consists of a CiteULike article ID, username, date, and time an article was added, as well as tags applied to the article. We used the article ID to crawl the CiteULike website, and we randomly downloaded 554,023 metadata files. These files contained more details about each article, including a link to the publisher's website, a list of authors, an abstract, a DOI, BibTex bibliographic information, the venue name, the year of publication, and a list of the users who had added that article to their personal article collections. We used an XML parser to clean and extract information from the CiteULike files. Using the BibTex field, we selected only the files that included either conference or journal data. Our final dataset contained 407,038 files. We then extracted the venue details from each article and collected a total of 1,317,336 postings of researcher–article pairs, as well as 614,361 researcher–venue pairs (Alhoori & Furuta, 2013). Fig. 1 shows the PVR recommendation system architecture.

We implemented user-based collaborative filtering (CF), item-based CF, stochastic gradient descent (SGD), and singular value decomposition (SVD++) algorithms from Apache Mahout (Apache Software Foundation, 2015) to recommend venues to researchers. We compared researchers with similar interests in terms of their PVRs. In recommendation systems, we need to find similarities between users or items. We choose to apply cosine similarity, Pearson correlation similarity, and Euclidean distance similarity in this study, because they are widely used in this type of analysis (Herlocker, Konstan, Borchers, & Riedl, 1999). The cosine similarity between two vectors is the angle between them and is usually useful for sparse data. The smaller the angle is to zero then the more similar those two vectors are. The Pearson correlation shows that when a series of ratings increase or decrease together. It is considered a centered version of cosine similarity (e.g., a cosine similarity when the two vectors have a mean of zero). The Euclidean distance similarity utilizes the distance between two vectors to calculate the distance between users.

The cosine similarity ($sim_{x,u}$) between a researcher x and another researcher u was computed as shown in Eq. (2), where $\vec{x}$ and $\vec{u}$ are two vectors representing the ratings of the two researchers, $||\vec{u}||$ is the vector's Euclidian length, and n is the number of venues rated by both researchers. The cosine similarity is the cosine angle between them.

$$sim_{x,u} = \cos(\theta) = \frac{\vec{x} \cdot \vec{u}}{||\vec{x}|| \times ||\vec{u}||} = \frac{\sum_{v=1}^{n}(r_{x,v})(r_{u,v})}{\sqrt{\sum_{v=1}^{n}(r_{x,v})^2}\ \sqrt{\sum_{v=1}^{n}(r_{u,v})^2}} \tag{2}$$

The Pearson correlation similarity ($sim_{x,u}$) is measured by Eq. (3). $\bar{r}_u$ is the average PVR for researcher u.

$$sim_{x,u} = \frac{\sum_{v=1}^{n}\left(r_{x,v} - \bar{r}_x\right)\left(r_{u,v} - \bar{r}_u\right)}{\sqrt{\sum_{v=1}^{n}\left(r_{x,v} - \bar{r}_x\right)^2}\ \sqrt{\sum_{v=1}^{n}\left(r_{u,v} - \bar{r}_u\right)^2}} \tag{3}$$

Eq. (4) shows the Euclidean distance. $V_{x,u}$ is the set of venues rated by both x and u.

$$Euclidean\ distance(x,u) = \sqrt{\frac{\sum_{v \in V_{x,u}}(r_{x,v} - r_{u,v})^2}{|V_{x,u}|}} \tag{4}$$

In the Euclidian distance similarity algorithm, a greater distance indicates fewer similar researchers. We therefore used $(1/(1 + distance)$ to identify similar researchers. Active researchers would have many similar venues with other researchers, which would create high correlations based on just a few co-rated venues. Therefore, for researchers who shared fewer co-rated venues than a threshold, we applied significance weighting (Herlocker et al., 1999), which reduced the overestimated similarity weight by a factor proportional to the number of co-rated venues. However, for researchers who shared more venues than the threshold, we did not adjust the similarity weight, as the more venues that are shared the more reliable the similarity. We found that this technique improved the accuracy of our results.

Users tend to assign different ranges of ratings. In other words, some users may generally assign high ratings whereas others may generally assign low ratings. Therefore, we normalized the ratings using a user mean-centering prediction. Prediction $p_{x,v}$ for an active user x and for venue v is measured by Eq. (5). $\bar{r}_x$ is the average rating assigned by user x to all the rated items. $U_v(x)$ is the set of user x's neighbors (similar users) who rated venue v. $\bar{r}_u$ is the average rating for user u for the items rated by both x and u (i.e., all the co-rated items).

$$p_{x,v} = \bar{r}_x + \frac{\sum_{u \in U_v(x)}\left(r_{u,v} - \bar{r}_u\right)sim_{x,u}}{\sum_{u \in U_v(x)}|sim_{x,u}|} \tag{5}$$

We also calculated the item mean-centering prediction, as shown in Eq. (6). $\bar{r}_v$ is the average rating of venue v for all users. $W_x(v)$ is the set of venues similar to venue v and rated by user x (venues rated by x as most similar to v). $\bar{r}_w$ is the average rating for venue w derived from the ratings of all the users who rated venues w and v.

$$p_{x,v} = \bar{r}_v + \frac{\sum_{w \in W_x(v)}\left(r_{x,w} - \bar{r}_w\right)sim_{v,w}}{\sum_{w \in W_x(v)}|sim_{v,w}|} \tag{6}$$

4.2. Evaluation metrics

We used a Boolean recommendation as a baseline and compared it with recommendations for scholarly venues based on PVR implicit ratings. Boolean ratings assume that all venues added by researchers are good venues and receive the highest rating. In the case of Boolean ratings, we used the log-likelihood similarity algorithm (Dunning, 1993). To rank the Boolean recommendations, venues affiliated with a large number of similar users were weighted more heavily (Owen, Anil, Dunning, & Friedman, 2011).

To measure the recommendations' performance, we measured precision, recall, and normalized discounted cumulative gain (NDCG) (Järvelin & Kekäläinen, 2002; McNee, Riedl, & Konstan, 2006). Precision is derived by dividing the number of relevant venues recommended according to the researcher's interests by the number of recommended venues, as shown in Eq. (7). Recall is derived by dividing the number of relevant venues recommended by the number of relevant venues, as shown in Eq. (8). For each user, the top 10 venues ranked by PVR were removed and the percentage of those 10 venues that appeared in the proposed top recommendations constituted the precision at 10 (P@10).

$$Precision = \frac{|relevant\ venues \cap top\ venues|}{|top\ venues|} \tag{7}$$

$$Recall = \frac{|relevant\ venues \cap\ top\ venues|}{|relevant\ venues|} \tag{8}$$

Discounted cumulative gain (DCG) measures the extent to which a venue ranking is relevant to a user's ideal ranking, as shown in Eq. (9).

$$DCG_p = \sum_{v=1}^{p} \frac{2^{rel_v} - 1}{log_2 (1 + v)} \tag{9}$$

rel_v is the relevance assigned by a researcher to the venue at position p. We measured the normalized discounted cumulative gain (NDCG), which ranges from 0.0 to 1.0, with 1.0 as the ideal ranking, as shown in Eq. (10). As recommendation lists vary in length, we used NDCG. $IDCG_p$ is the maximum possible ideal DCG at position p.

$$NDCG_p = \frac{DCG_p}{IDCG_p} \tag{10}$$

We also incorporated user coverage (Good et al., 1999; Herlocker, Konstan, Terveen, & Riedl, 2004; Sarwar et al., 1998), which is the percentage of users for whom the system was able to recommend venues. Additionally, we tested for the normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE), which are independent rating scales. MAE (Shani & Gunawardana, 2009), the absolute deviation of a researcher's predicted PVR and observed PVR, is calculated as shown in Eq. (11). $p_{u,v}$ is the predicted rating for venue v, and $r_{u,v}$ is the actual rating.

$$MAE = \frac{\sum_{v=1}^{n} |p_{u,v} - r_{u,v}|}{n} \tag{11}$$

RMSE is measured using the square root of the average squared difference between a researcher's predicted PVR and observed PVR as shown in Eq. (12).

$$RMSE = \sqrt{\frac{\sum_{v=1}^{n} (p_{u,v} - r_{u,v})^2}{n}} \tag{12}$$

We used 70% of the data as a training set and 30% as a test set. We selected recommendations by choosing a threshold per user that was equal to the user's average PVR.

5. Results and discussion

We began by comparing user similarities with and without significance weighting. Collaborative filtering is affected by the cold-start problem (Schein, Popescul, Ungar, & Pennock, 2002) in which the system cannot produce good recommendations for new researchers or unrated venues. As our dataset includes thousands of venues and as each researcher would only know few of them, many of the venues will not have a rating from any given researcher. In addition, the Pearson correlation was not able to compute any similarity between users in some cases. For example, researchers who had added only one reference to their libraries and no other researchers share the same or similar reference. Therefore, venues that did not have an indirect rating were inferred to have an average rating.

Fig. 2 shows that the use of significance weighting improved the accuracy, recall, and NDCG. The use of inferred ratings showed some improvement in the results as the neighborhood size increased. Further, Fig. 2 also shows that Euclidean-weighting achieved better precision, recall, NDCG, and users' coverage than Pearson-weighting. We tested other similarities such as Spearman correlation similarity and Tanimoto coefficient similarity, but we found that Pearson correlation similarity and Euclidean distance similarity achieved better results.

We then compared similarities that used PVR ratings and the user-based collaborative filtering algorithm with the baseline Boolean recommendation, as shown in Fig. 3. Fig. 3(a–c) demonstrates that in general the PVR implicit ratings achieved higher precision, recall, and NCDG at lower neighborhood sizes. This result shows that in multidisciplinary environments such as CiteULike, researchers with a high level of similarity tend to cluster in small groups, which explains why the baseline achieved better results when the number of similar researchers increased (neighborhood size). Fig. 3(d) shows the users' coverage and that the PVR model was able to provide recommendations for up to 98% of users.

Fig. 4 illustrates the use of thresholds for researchers who are at least T percentage similar instead of fixed neighborhood sizes. Pearson-weighting achieved the highest P@10 and the highest NDCG, whereas the Boolean recommendations achieved the highest recall and the highest coverage at low thresholds. Fig. 4(a) and (c) shows that the threshold increases and reaches its maximum at a value around 0.6. Then it starts to drop since the neighborhood does contain fewer researchers.

We measured NMAE and NRMSE at different neighborhood sizes as Fig. 5 shows, and found that the Euclidean-weighting achieved the lowest NMAE as well as the lowest NRMSE. Fig. 5 also shows that Pearson-weighting and cosine similarity result in the highest error. Figs. 2, 3, and 5 show that Euclidean-weighting achieved the highest level of accuracy and the lowest level of errors.
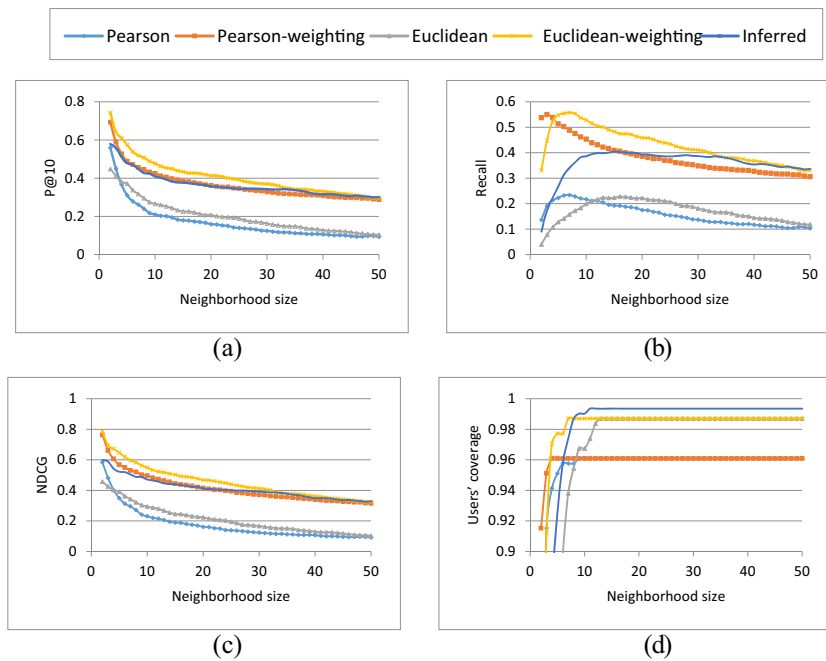
Fig. 2. Comparison of the user-based CF algorithm with different similarities and neighborhood sizes.
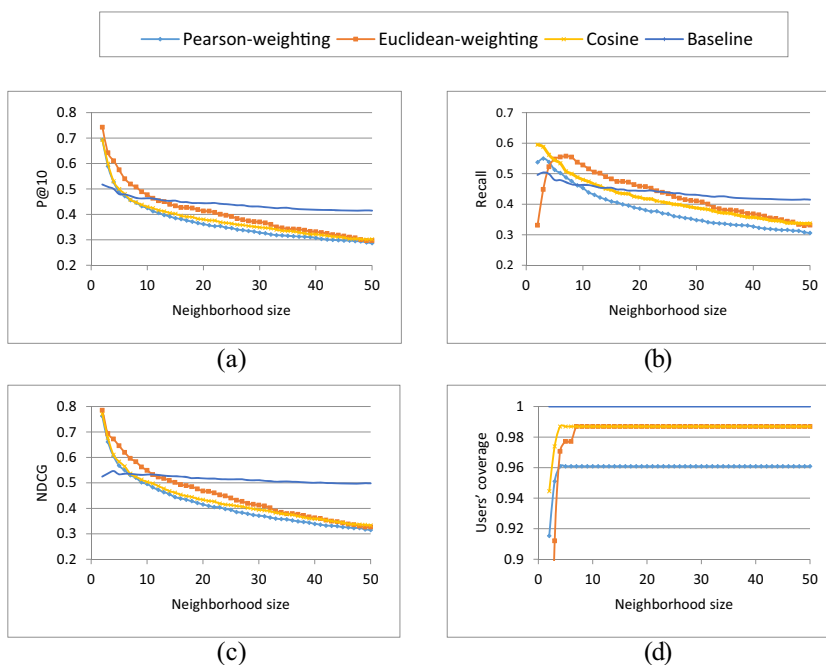

Fig. 3. Comparison of the user-based CF algorithm with similarities that use PVR ratings and the baseline at different neighborhood sizes.

We compared the performance of four algorithms that used PVR ratings at different percentages of the training set (Fig. 6), and we found that SVD++ achieved the lowest NMAE and the lowest NRMSE. Item-based CF performed better than User-based CF, which shows that researchers' interests and implicit ratings are changing over time.

We tested another model and it showed no improvements during evaluation. In that model, we took each user's references to a specific venue in a single year and compared those references to the total added references for that user that year. This method would show the importance of a particular venue to a user in a year, but we found that other factors, such as venue size, dominated in practice.
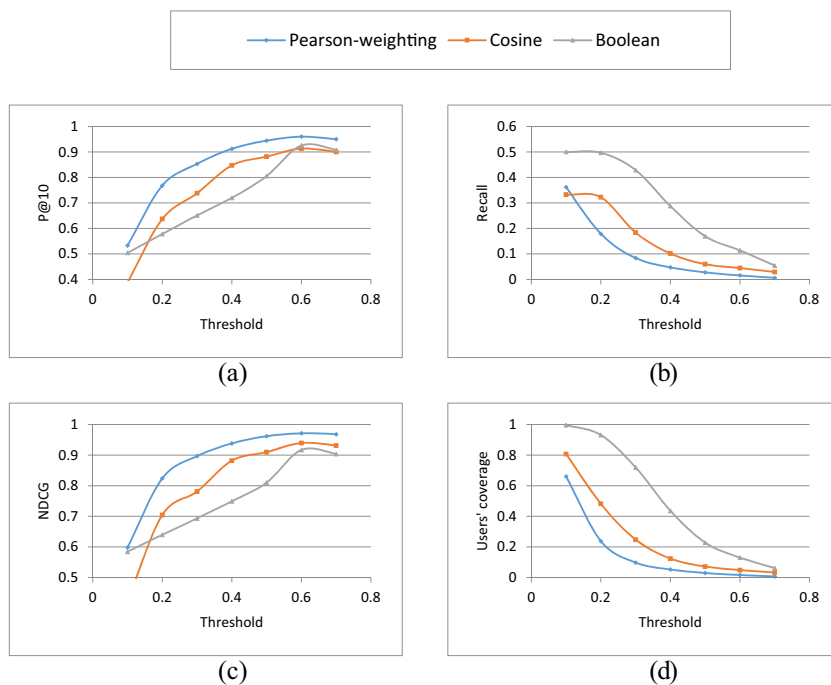
Fig. 4. Comparison of user-based CF performance using different similarities and thresholds.
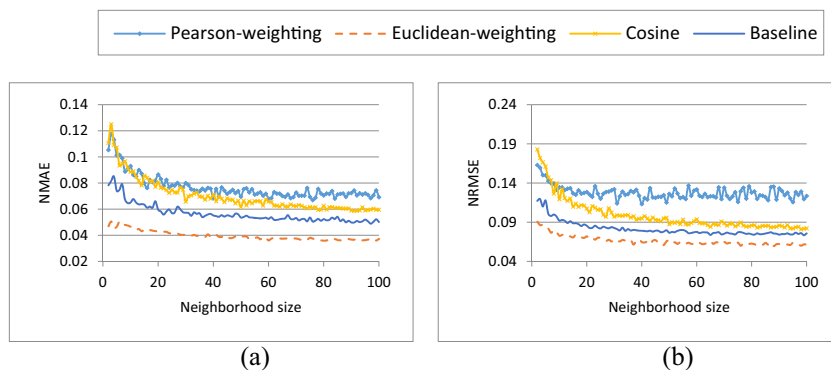

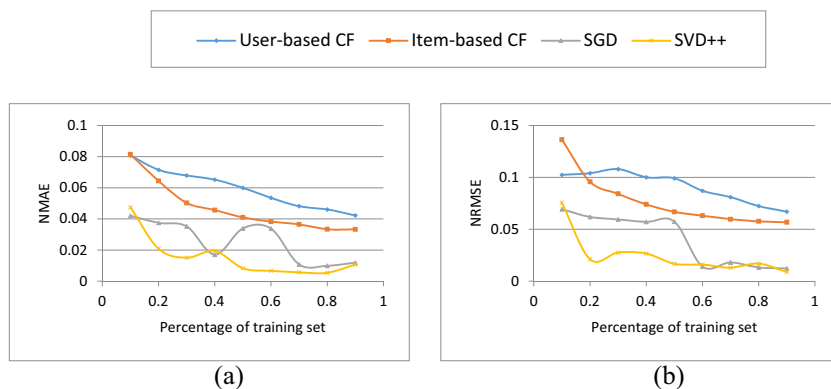**Fig. 5.** NMAE and NRMSE for user-based CF with different neighborhood sizes.

Fig. 6. Performance comparison of different recommendation algorithms at different training ratios.

The use of explicit data, such as favorites or ratings for references, could improve the accuracy of recommendations. Explicit data of this nature show how interested a researcher is in an article in a much less ambiguous way. In this regard, CiteULike provides two optional but important fields that can affect venue ratings. The first field is a researcher's explicit rating of an article, and the second field is the priority a researcher has assigned to reading an article. These explicit ratings would improve PVR measurements, especially for researchers who have an interest in smaller-size venues. However, in order to collect data pertaining to these two fields, it would be necessary to construct a new dataset. Our current dataset contains unique article IDs, rated only by the first researcher who added the article to CiteULike.

6. Conclusion and future work

Multidisciplinary research areas are growing at a tremendous rate, and the number of scholarly venues is increasing every year. Researchers need to discover venues that are of interest to them, and research institutions need to be aware of these venues. In this paper, using data from an academic social network, we described an approach to recommend scholarly venues for researchers to follow and/or to publish their work in based on their current interests.

We developed a new weighting strategy for rating venues based not only on personal references, but also on the temporal factor of when the references were added. Of the similarity metrics, Euclidean-weighting achieved the best precision, recall, and NDCG. Of the recommendation algorithms, SVD++ achieved the lowest error rates. Our experiments with this strategy using a real dataset produced results that showed improvements in accuracy and ranking quality compared with a standard baseline. A number of factors will be investigated to improve the results and recommendation quality, including the total number of papers published in a venue, the number of online references to a venue in an academic social network, the average number of references added by researchers to an online reference management system, the dates on which references were added to the researchers' repositories, and the readership statistics for an article.

In future research, we plan to enhance the quality of our generated recommendations by using a researcher's trustworthiness and reputation (Alhoori, Alvarez, Furuta, Miguel Mu, & Urbina, 2009), cited references (Thor, Marx, Leydesdorff, & Bornmann, 2016), and various altmetrics (Thelwall, Haustein, Larivière, & Sugimoto, 2013) with the goal of improving accuracy, diversity, novelty, and serendipity (Ge, Delgado-Battenfeld, & Jannach, 2010). Also planned is a user study through which we will collect explicit ratings to compare with our implicit ratings. The system will begin similarly, using metadata of articles, such as title, abstract, keywords, and tags, to recommend venues, but will diverge into an analysis of explicit user-provided ratings. These experiments will use a hybrid approach implementing both collaborative filtering and content-based filtering. In addition, other factors that affect researchers' choices will be considered, such as budget availability and the ability to travel in cases such as conferences or workshops.

Author contributions

Hamed Alhoori: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.

Richard Furuta: Conceived and designed the analysis; Wrote the paper.

Acknowledgements

References

Alhoori, H., & Furuta, R. (2011). Understanding the dynamic scholarly research needs and behavior as applied to social reference management. *Proceedings of the 15th international conference on theory and practice of digital libraries, Vol. 6966*, 169–178. http://dx.doi.org/10.1007/978-3-642-24469-8_19

Alhoori, H., & Furuta, R. (2013). *Can social reference management systems predict a ranking of scholarly venues?* pp. 138–143. *Proceedings of the 17th international conference on theory and practice of digital libraries* (Vol. 8092) Springer. http://dx.doi.org/10.1007/978-3-642-40501-3_14

Alhoori, H., Alvarez, O., Furuta, R., Miguel Mu, N., & Urbina, E. (2009). *Supporting the creation of scholarly bibliographies by communities through online reputation based social collaboration*. pp. 180–191. *Proceedings of the 13th European conference on research and advanced technology for digital libraries* (Vol. 5714) Corfu, Greece: Springer. http://dx.doi.org/10.1007/978-3-642-04346-8_19

Alhoori, H. (2016). How to identify specialized research communities related to a researcher's changing interests. *Proceedings of the 16th ACM/IEEE-CS on joint conference on digital libraries – JCDL '16*, 239–240. http://dx.doi.org/10.1145/2910896.2925450

Apache Software Foundation. (2015). *Apache Mahout.*. Retrieved from http://mahout.apache.org/

Basu, C., Cohen, W. W., Hirsh, H., & Nevill-Manning, C. (2001). Technical paper recommendation: A study in combining multiple information sources. *Journal of Artificial Intelligence Research*, *14*(1), 231–252. Retrieved from http://arxiv.org/abs/1106.0248

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2015). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, 1–34.

Blake, C., & Pratt, W. (2006). Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology*, *57*(13), 1740–1749. http://dx.doi.org/10.1002/asi

Bottou, L., & Bousquet, O. (2008). The tradeoffs of large scale learning. *Advances in Neural Information Processing Systems* (Vol. 20, pp. 161–168). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.5389&;rep=rep1&type=pdf.

Boukhris, I., & Ayachi, R. (2014). A novel personalized academic venue hybrid recommender. *IEEE 15th international symposium on computational intelligence and informatics*, 465–470. http://dx.doi.org/10.1109/CINTI.2014.7028720

Buchanan, G., Cunningham, S. J., Blandford, A., Rimmer, J., & Warwick, C. (2005). *Information seeking by humanities scholars*. pp. 218–229. *Proceedings of the 9th European conference on digital libraries* (Vol. 3652) Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/11551362_20

Caragea, C., Silvescu, A., Mitra, P., & Giles, C. L. (2013). Can't see the forest for the trees? A citation recommendation system. *Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries*, 111–114. http://dx.doi.org/10.1145/2467696.2467743

Chu, S. K.-W., & Law, N. (2007). Development of information search expertise: Postgraduates' knowledge of searching skills. *Portal: Libraries and the Academy*, *7*(3), 295–316. http://dx.doi.org/10.1353/pla.2007.0028

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74. Retrieved from http://dl.acm.org/citation.cfm?id=972450.972454

Errami, M., Wren, J. D., Hicks, J. M., & Garner, H. R. (2007). eTBLAST: A web server to identify expert reviewers, appropriate journals and similar publications. *Nucleic Acids Research*, *35*(Web Server), W12–W15. http://dx.doi.org/10.1093/nar/gkm221

Farooq, U., Song, Y., Carroll, J. M., & Giles, C. L. (2007). Social bookmarking for scholarly digital libraries. *IEEE Internet Computing*, *11*(6), 29–35. http://dx.doi.org/10.1109/MIC.2007.135

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*. pp. 257–260. New York, USA: ACM Press. http://dx.doi.org/10.1145/1864708.1864761

Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J., et al. (1999). Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the sixteenth national conference on artificial intelligence and the eleventh innovative applications of artificial intelligence conference innovative applications of artificial intelligence*. pp. 439–446. JOHN WILEY & SONS LTD. http://dx.doi.org/10.1016/j.tet.2006.06.092

Gove, R., Dunne, C., Shneiderman, B., Klavans, J., & Dorr, B. (2011). Evaluating visual and statistical exploration of scientific literature networks. In *IEEE symposium on visual languages and human-centric computing (VL/HCC)* (pp. 217–224). http://dx.doi.org/10.1109/VLHCC

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 230–237). http://dx.doi.org/10.1145/312624.312682

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53. http://dx.doi.org/10.1145/963770.963772

Huynh, T., & Hoang, K. (2012). Modeling collaborative knowledge of publishing activities for research recommendation. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*,. Vol. 7653 LNAI, pp. 41–50.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, *20*(4), 422–446. http://dx.doi.org/10.1145/582415.582418

Küçüktunç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2012). *Recommendation on academic networks using direction aware citation analysis. arXiv:1205.1143.*. Retrieved from http://arxiv.org/abs/1205.1143

Kang, N., Doornenbal, M. A., & Schijvenaars, R. J. A. (2015). Elsevier journal finder: Recommending journals for your paper. In *Proceedings of the 9th ACM conference on recommender systems – RecSys '15*. pp. 261–264. New York, USA: ACM Press. http://dx.doi.org/10.1145/2792838.2799663

Khrouf, H., & Troncy, R. (2013). Hybrid event recommendation using linked data and user diversity. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 185–192). Retrieved from http://dl.acm.org/citation.cfm?doid=2507157.2507171

Klamma, R., Cuong, P. M., & Cao, Y. (2009). You never walk alone: Recommending academic events based on social network analysis. *Proceedings of the first international conference on complex science, Vol. 4*, 657–670. Retrieved from http://www.springerlink.com/index/v1u803162565667l7.pdf

Kochen, M., & Tagliacozzo, R. (1974). Matching authors and readers of scientific papers. *Information Storage and Retrieval*, *10*(5–6), 197–210.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30–37. http://dx.doi.org/10.1109/MC.2009.263

Kucuktunc, O., Saule, E., Kaya, K., & Catalyurek, U. V. (2013). TheAdvisor: A webservice for academic recommendation. *Proceedings of the ACM/IEEE joint conference on digital libraries*, 433–434. Retrieved from http://dx.doi.org/10.1145/2467696.2467752

Kuhn, M., & Wattenhofer, R. (2008). The layered world of scientific conferences. *Progress in WWW Research and Development* (pp. 81–92). 10.1007/978-3-540-78849-2_10.

Kuruppu, P. U., & Gruber, A. M. (2006). Understanding the information needs of academic scholars in agricultural and biological sciences. *The Journal of Academic Librarianship*, *32*(6), 609–623. http://dx.doi.org/10.1016/j.acalib.2006.08.001

Lam, S. K., & Riedl, J. (2004). Shilling recommender systems for fun and profit. In *Proceedings of the 13th conference on world wide web*. pp. 393–402. New York, USA: ACM Press. http://dx.doi.org/10.1145/988672.988726

Lu, Z., Xie, N., & Wilbur, W. J. (2009). Identifying related journals through log analysis. *Bioinformatics*, *25*(22), 3038–3039.

Luong, H.P., Huynh, T., Gauch, S., & Hoang, K. (2012). Exploiting Social Networks for Publication Venue Recommendations. *KDIR* (pp. 239–245).

Luong, H., Huynh, T., Gauch, S., Do, P., & Hoang, K. (2012). Publication venue recommendation using author network's publication history. *Proceedings on the 4th Asian conference on intelligent information and database systems*, 426–435. http://dx.doi.org/10.1007/978-3-642-28493-9

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. In G. M. Olson, & R. Jeffries (Eds.), *CHI '06 extended abstracts on human factors in computing systems* (pp. 1097–1101). ACM. http://dx.doi.org/10.1145/1125451.1125659

Medvet, E., Bartoli, A., & Piccinin, G. (2014). Publication venue recommendation based on paper abstract. In *Proceedings of the IEEE 26th international conference on tools with artificial intelligence*. pp. 1004–1010. IEEE. http://dx.doi.org/10.1109/ICTAI

Minkov, E., Charrow, B., Ledlie, J., Teller, S., & Jaakkola, T. (2010). Collaborative future event recommendation. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. pp. 819–828. New York, USA: ACM Press. http://dx.doi.org/10.1145/1871437.1871542

Murphy, J. (2003). Information-seeking habits of environmental scientists: A study of interdisciplinary scientists at the environmental protection agency in research triangle park, North Carolina. *Issues in Science and Technology Librarianship*, 38.

Owen, S., Anil, R., Dunning, T., & Friedman, E. (2011). *Mahout in action*. Manning Publications.

Perlman, G. (1991). The HCI bibliography project. *ACM SIGCHI Bulletin − Special Issue: Computer Supported Cooperative Work*, *23*(3), 15–20. http://dx.doi.org/10.1145/126505.126507

Pham, M. C., Cao, Y., Klamma, R., & Jarke, M. (2011). A clustering approach for collaborative filtering recommendation using social network analysis. *Journal Of Universal Computer Science*, *17*(4), 583–604. Retrieved from http://www.jucs.org/

Protasiewicz, J., Pedrycz, W., Kozłowski, M., Dadas, S., Stanisławek, T., Kopacz, A., et al. (2016). A recommender system of reviewers and experts in reviewing problems. *Knowledge-Based Systems*, *106*, 164–178. http://dx.doi.org/10.1016/j.knosys.2016.05.041

Quercia, D., Lathia, N., Calabrese, F., Di Lorenzo, G., & Crowcroft, J. (2010). Recommending social events from mobile phone location data. In *Proceedings of the IEEE 10th international conference on data mining* (pp. 971–976). http://dx.doi.org/10.1109/ICDM. 2010.152

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58. http://dx.doi.org/10.1145/245108.245121

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In R. Furuta, & C. M. Neuwirth (Eds.), *Proceedings of the ACM conference on Computer supported cooperative work* (pp. 175–186). New York, USA: ACM Press. http://dx.doi.org/10.1145/192844.192905

Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J., Miller, B., & Riedl, J. (1998). Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In *Proceedings of the ACM conference on Computer supported cooperative work* (pp. 345–354). http://dx.doi.org/10.1145/289444.289509

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Application of dimensionality reduction in recommender system – a case study. *ACM WebKDD web mining for e-commerce workshop.*

Sarwar, B., Karypis, G., Konstan, J., & Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on world wide web* (pp. 285–295). http://dx.doi.org/10.1145/371920.372071

Schafer, J., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. *The Adaptive Web* (pp. 291–324). doi:10.1007/978-3-540-72079-9_9.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 253–260). http://dx.doi.org/10.1145/564376.564421

Schuemie, M. J., & Kors, J. A. (2008). Jane: Suggesting journals, finding experts. *Bioinformatics*, 24(5), 727–728. http://dx.doi.org/10.1093/bioinformatics/btn006

Shani, G., & Gunawardana, A. (2009). Evaluating recommendation systems. In *Recommender systems handbook.* pp. 257–297. Springer. http://dx.doi.org/10.1007/978-0-387-85820-3_8

Shenk, D. (1997). *Data smog: Surviving the information glut. new York times on the web.* New York, NY: HarperCollins.

Silva, T., Ma, J., Yang, C., & Liang, H. (2015). A profile-boosted research analytics framework to recommend journals for manuscripts. *Journal of the Association for Information Science and Technology*, 66(1), 180–200. http://dx.doi.org/10.1002/asi.23150

Singh, M., Chakraborty, T., Mukherjee, A., & Goyal, P. (2016). Is this conference a top-tier? ConfAssist: An assistive conflict resolution framework for conference categorization. *Journal of Informetrics*, 10(4), 1005–1022. http://dx.doi.org/10.1016/j.joi.2016.08.001

Song, Y., Zhang, L., & Giles, C. L. (2011). Automatic tag recommendation algorithms for social recommender systems. *ACM Transactions on the Web*, 5(1), 1–31. http://dx.doi.org/10.1145/1921591.1921595

Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2), 337–360. http://dx.doi.org/10.1111/j.1540-5915.1999.tb01613.x

Springer. (2015). *Journals subscription, recommend to your librarian/information specialist..* Retrieved from http://www.springer.com/generic/order/journals+subscription?SGWID=0-40514-0-0-0

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One*, 8(5), e64841. http://dx.doi.org/10.1371/journal.pone.0064841

Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization. *Journal of Informetrics*, 10(2), 503–515. http://dx.doi.org/10.1016/j.joi.2016.02.005

Wu, I. C., & Niu, Y. F. (2015). Effects of anchoring process under preference stabilities for interactive movie recommendations. *Journal of the Association for Information Science and Technology*, 66(8), 1673–1695. http://dx.doi.org/10.1002/asi.23280

Xia, F., Asabere, N. Y., Rodrigues, J. J. P. C., Basso, F., Deonauth, N., & Wang, W. (2013). Socially-aware venue recommendation for conference participants. In *2013 IEEE 10th international conference on ubiquitous intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing* (pp. 134–141). http://dx.doi.org/10.1109/UIC-ATC

Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics, 8*(2), 295–309.

Yang, Z., & Davison, B. D. (2012). Venue recommendation: Submitting your paper with style. *Proceedings of the 11th international conference on machine learning and applications* (Vol. 1) (pp. 681–686). http://dx.doi.org/10.1109/ICMLA. 2012.127