MINING, MODELING, AND LEVERAGING MULTIDIMENSIONAL WEB METRICS TO SUPPORT SCHOLARLY COMMUNITIES

A Dissertation

by

HAMED MOHAMED HASAN TAQI MOHAMED ALHOORI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Richard Furuta
Committee Members,	Frank Shipman
	James Caverlee
	Lauren Cifuentes
Head of Department,	Dilma Da Silva

August 2015

Major Subject: Computer Science

ABSTRACT

The significant proliferation of scholarly output and the emergence of multidisciplinary research areas are rendering the research environment increasingly complex. In addition, an increasing number of researchers are using academic social networks to discover and store scholarly content. The spread of scientific discourse and research activities across the web, especially on social media platforms, suggests that far-reaching changes are taking place in scholarly communication and the geography of science.

This dissertation provides integrated techniques and methods designed to address the information overload problem facing scholarly environments and to enhance the research process. There are four main contributions in this dissertation. First, this study identifies, quantifies, and analyzes international researchers' dynamic scholarly information behaviors, activities, and needs, especially after the emergence of social media platforms. The findings based on qualitative and quantitative analysis report new scholarly patterns and reveals differences between researchers according to academic status and discipline.

Second, this study mines massive scholarly datasets, models diverse multidimensional non-traditional web-based indicators (altmetrics), and evaluates and predicts scholarly and societal impact at various levels. The results address some of the limitations of traditional citation-based metrics and broaden the understanding and utilization of altmetrics. Third, this study recommends scholarly venues semantically

related to researchers' current interests. The results provide important up-to-the-minute signals that represent a closer reflection of research interests than post-publication usage-based metrics.

Finally, this study develops a new scholarly framework by supporting the construction of online scholarly communities and bibliographies through reputation-based social collaboration, through the introduction of a collaborative, self-promoting system for users to advance their participation through analysis of the quality, timeliness and quantity of contributions. The framework improves the precision and quality of social reference management systems.

By analyzing and modeling digital footprints, this dissertation provides a basis for tracking and documenting the impact of scholarship using new models that are more akin to reading breaking news than to watching a historical documentary made several years after the events it describes.

DEDICATION

To my family

ACKNOWLEDGEMENTS

First and foremost, I thank the Almighty Creator, the most gracious and the most merciful, for his guidance and countless blessings throughout my life and during my higher education studies. It is only with his care and love and the power and knowledge he has bestowed on me that I have succeeded in reaching my goals. There are no words to describe what his presence in my life means to me.

I am forever grateful to my amazing family for always supporting my endeavors, for their continuous encouragement since my childhood, for planting and growing the seeds of curiosity, inspiration, confidence, and hope, and for their support in so many ways, despite the thousands of miles between us. I will never forget the true meaning of unconditional love as I have learned it from them throughout the years.

I would like to express my sincere gratitude to my advisor Dr. Richard Furuta for his patience, guidance, and constructive feedback. I have learned much from him, both in terms of academic inquiry and life lessons. For instance, I learned that buzzwords are not necessarily a good basis for research, that moving with the crowd can be useless, and that I should take the risk of blazing new trails in untested intellectual territories. I am thankful to my advisor for giving me the intellectual freedom to learn, discover, and grow while ensuring that I stay on the right track. I am grateful that he encouraged me as I explored my ideas and supported my presentations at prestigious international conferences and that he taught me how to become an independent researcher.

I would like to thank both my family and my advisor for the many kindnesses they have shown me and the many things they have taught me: I have learned that becoming motivated and then motivating others to try out new approaches can be a recipe for change and that encouraging and helping others to find their own passion and build their own practices can be a recipe for progress and fulfillment. It was in these ways that I discovered my own real interests, connected me to understanding the beauty of research and knowledge, became a lifelong learner and educator, and found and remembered the purpose of life.

I am thankful to my dissertation committee Dr. Frank Shipman, Dr. James Caverlee, and Dr. Lauren Cifuentes for their thoughtful insight and advice. Thanks, too, to my colleagues in other studies for their assistance and for our discussions relating to this work. I wish to offer my particular thanks to Carole Thompson who took the time to review my dissertation and for her valuable comments and suggestions. To all my teachers and professors, thank you for preparing me to compete and collaborate with the world's top researchers.

I must also express my thanks to the ELISQ project that supported me during the last three years of my doctoral program. This dissertation was made possible in part by NPRP grant # 4-029-1-007 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the author. A special note of thanks also goes to the ELISQ project members with whom I have so enjoyed working. My heartfelt thanks also go to Dr. Edward Fox who played an important role in facilitating my research. Thanks to University of Bahrain for sponsoring me during my master's program and during the initial years of my Ph.D. program.

I have been very fortunate, too, to receive various scholarships and other awards from Texas A&M University and other entities, which provided me with both support and encouragement during my studies.

My thanks go just as much to all the friends and colleagues at Texas A&M University and beyond, whom I have come to know during my years of study. Thanks to the Aggie community who provided a second home, wonderful moments and memorable experiences.

Many people have provided me with guidance, support, and motivation along this journey. I am thankful to all of them.

NOMENCLATURE

5-IF	Five-year Impact Factor
ARR	Active Researchers Rating
BDL	Bibliographic Digital Library
CF	Collaborative Filtering
CIBO	Cervantes International Bibliography Online
DOI	Digital Object Identifier
GDP	Gross Domestic Product
GERD	Gross Domestic Expenditure on Research and Development
IF	Impact Factor
ISSN	International Standard Serial Number
JCR	Journal Citation Reports
JSI	Journal Social Impact
NOA	Non-Open Access
NSF	National Science Foundation
OA	Open Access
OAAA	Open Access Altmetric Advantage
ORSC	Online Reputation-based Social Collaboration
PVR	Personal Venue Rating
RCAR	Research Community Article Rating
SRM	Social Reference Management

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation

Billions of dollars are spent each year on research and the resulting publications [1]. However, research outcomes are rarely leveraged to the fullest extent possible. This can be attributed to the fact that scholarly communities face multiple challenges. On this point, former director of the National Library of Medicine, Martin M. Cummings, summed up the situation like this: “Can a productive scientist keep abreast of a scientific literature that doubles in size every fifteen years and shows evidence of continued exponential growth during this decade? I believe that it is no longer possible to do so, even in a limited field or discipline” [2]. The massive increase in materials available to scholars has rendered the scholarly research environment correspondingly more complex, even though researchers continue to read more articles [3][4]. Harnad et al. [5] estimate that nearly 2.5 million articles are published yearly. Khabsa and Giles [6] estimate that at least 114 million English-language research documents are accessible on the web, of which 24% are freely available. Attempts to quickly find literature related to any given topic among millions of research documents can be similar to finding a needle in a haystack. As a result, editors and reviewers criticize manuscripts submitted for publication that fail to locate, analyze, and synthesize related scholarly work [7].

Not only is the number of scholarly publications increasing, but also there is the additional complication presented by a research landscape that is becoming less compartmentalized. There are, for example, increasingly complex academic sub-disciplines and emerging interdisciplinary research areas, events, and venues (e.g., journals, conferences, symposiums, workshops, and seminars). In this competitive and sophisticated research environment, it is challenging for researchers to remain up-to-date with new findings, even within their own disciplines [8][9]. Interdisciplinary researchers are required to possess substantial knowledge, a broad vocabulary, and the competence to pursue numerous research approaches in two or more research areas [9]. For example, a computer scientist who is interested in the field of neuroscience but unfamiliar with its venues and current ongoing research is likely to experience challenges finding related work. Further, in scholarly communities, "context-drift" is becoming popular as researchers expand, evolve, or adapt their interests in rapidly changing subject areas over time.

Previous studies have used several approaches to filtering intellectual resources, including peer review and citation analysis, each of which has benefits and limitations. Moreover, many researchers, departments, and research communities act unilaterally as if they were isolated islands: that is, they define problems, solve them using limited methods and theories, and then share the results with a small number of community members. Yet, the scale of many problems faced by researchers requires interdisciplinary and cross-continental research.

In this epoch of big data, conducting comprehensive research in a fast-paced and interconnected world requires advanced technologies to assist researchers in discovering related research, establishing a thorough understanding of given problems and potential solutions, extracting and modeling data from numerous sources, visualizing patterns, and generating insights in order to tackle critical challenges. This dissertation addresses these goals by using human-centered computing and data science approaches.

This dissertation begins by identifying global researchers' emerging information behaviors, patterns, and needs. Next, it investigates and models various new web indicators from several dimensions and then continues by evaluating societal impact, predicting scholarly impact, and recommending scholarly venues. Finally, the dissertation develops new technologies to support scholarly communities and to leverage more scientific knowledge than is the case at present.

1.2. Overview of this Dissertation

This Subsection summarizes the main research problems and contributions of this dissertation as follows:

1. Given the abundance of scholarly products—especially in environments created by the advent of social networking services—little is known about international scholarly information needs, information-seeking behavior, and information use. Section 2 aims to address these gaps by conducting an in-depth analysis on researchers in two countries, learn about their research attitudes, practices, tactics, strategies, and expectations, as well as the obstacles faced during research endeavors. Based on this analysis, the study identifies and describes new behavior patterns on the part of researchers as they engage in the information-seeking process. The analysis reveals that the use of academic social networks has remarkable effects on various scholarly activities. Further, this study identifies differences between students and faculty members in regard to their use of academic social networks, and

it identifies differences between researchers according to discipline. The researchers who participated in the present study represent a range of disciplinary and cultural backgrounds. However, the study reports a number of similarities in terms of the researchers' scholarly activities. Finally, highlights of the study illuminate some of the implications for the design of research platforms. Establishing the ground truth in this section offers a basis for understanding the research community's problems and concerns and for addressing a number of these in subsequent sections.

2. Recently, non-traditional web-based indicators, known as altmetrics have been used to measure the impact of research activities in broader dimensions than traditional metrics are capable of measuring. In Section 3, a study of altmetrics at the article, journal, country, and access levels investigates whether the online attention received by research articles is related to scholarly impact and/or to other factors. The study used 14 data sources: Twitter, Facebook, CiteULike, Mendeley, F1000, blogs, mainstream news outlets, Google Plus, Pinterest, Reddit, sites running Stack Exchange (Q&A), Sina Weibo, peer review sites (PubPeer and Publons), and policy documents. A new metric is defined, *Journal Social Impact (JSI)*, which compares with diverse citation-based metrics and finds significant correlations. These findings indicate that online attention to scholarly articles relates to traditional journal rankings and favors journals with a long history of scholarly impact [10]. This study found that journal-level altmetrics have strong significant correlations among themselves, compared with the weak correlations among article-level altmetrics. Another important finding is that Mendeley and Twitter have the highest usage and

coverage of scholarly activities. For journal-level altmetrics, the findings showed that the readership of academic social networks have the highest correlations with citation-based metrics. On the country-level, the study found that altmetrics can support efforts to evaluate research impact for all the countries studied [11]. This study compared altmetrics with several traditional metrics, and significant relationships were found between country-level altmetrics and the number of publications, number of citations, h-index, and gross domestic expenditure on research and development (GERD). This study also found a significant yearly increase in the number of articles published between 2010 and 2014 that received altmetrics. And, finally, the relationship between the access approach to scholarly articles (i.e., Open Access (OA) and Non-Open Access (NOA)) and altmetrics was explored [12]. A new metric was defined: the *Open Access Altmetric Advantage (OAAA)*. The findings showed that OA articles received higher altmetrics than the NOA articles for eight of the fourteen data sources investigated. These findings deepen the overall understanding of altmetrics and provide a basis for validating them, thus opening a new door to research discovery and evaluation.

3. Valid measurements and accurate predictions of the impact of scholarly products increase researchers' awareness and assists stakeholders in evaluating research progress. Several approaches are used to evaluate and rank scholarly content, including expert surveys, citation-based metrics (e.g., impact factor, SCImago journal rank indicator, Eigenfactor score, and h-index), and usage-based metrics (e.g., downloads and views). Section 4 proposes techniques to measure the social

impact of research outcomes and to predict the research impact. We describe a new multi-dimensional model that can measure, in real-time, the impact of research, based on the research community article rating (RCAR) [13]. Secondly, we compare the performance of RCAR to those of both altmetrics and traditional citation analysis, showing that RCAR and altmetrics can quantify an early impact of articles, i.e., within just a few days of publication, which is long before articles usually receive any formal citations. We then propose an approach to predict venue ranking based on scholarly references from an academic social network. We investigate the relationship between ranking methods for scholarly venues that use traditional citation-based metrics and propose a set of social-based metrics, finding a statistically significant relationship between the two approaches in relation to a number of general rankings, research areas, and sub-disciplines, with disciplinary differences. These results suggest that academic social networks have the potential to provide an early indicator of the influence of scholarly venues while addressing some of the limitations of citation-based metrics [14].

4. The number of scholarly events and venues is increasing rapidly, and researchers need to identify those related to their work in order to draw on the published research and to share their own findings. Yet, there is no rating system to assist researchers in analyzing the venues most relevant to their interests, which often evolve over time. Therefore, as opportunities to share scholarly work proliferate, so researchers may find it correspondingly difficult to determine the best venues to follow and likewise the venues most appropriate for publishing their own research. Section 5

recommends scholarly venues that are rated in terms of their relevance to any given researcher's specific activities and interests. We collected our data from an academic social network and modeled researchers' scholarly behavior in order to propose a new and adaptive implicit rating technique for venues. We conducted experiments and found that the academic social network studied can effectively recommend scholarly venues and that the proposed rating outperforms the baseline venue recommendation.

5. Bibliographic digital libraries (BDLs) constitute a significant research resource, and in recent years they have started to move from closed to social platforms, of which the latter are more open and interactive in nature. However, in making this transition, BDLs have faced challenges (e.g., from spam) in regard to maintaining a high level of precision—i.e., the ratio of relevant references retrieved by searches. In Section 6, we describe a hybrid approach that uses online social collaboration and reputation-based social moderation to (1) reduce the cost and speed up the construction of scholarly bibliographies and (2) ensure that these bibliographies are more comprehensive and accurate than current scholarly bibliographies. We implemented selected social features for an established digital humanities bibliography and compared the results with a number of other bibliographies. Through our approach, we were able to build a scholarly bibliography that, compared to established approaches produced significantly improved precision outcomes [15]. Section 7 concludes with a summary of future research plans.

2. ANATOMY OF INTERNATIONAL SCHOLARLY INFORMATION BEHAVIOR PATTERNS IN THE WAKE OF SOCIAL MEDIA*

2.1. Introduction

Establishing an understanding of researchers' scholarly activities, including the paths they take in this regard, is vital to the discovery of new strategies and techniques whereby researchers can maximize their information gains. Further, a sound knowledge base pertaining to the patterns that govern these activities—which will be referred to as "scholarly information behavior"—would also facilitate the efforts of libraries, publishers, and other information providers to tailor services, develop specialized collections, and build academic digital libraries and research assessment tools [16].

Over the past decade, social networking services have been widely used in academia and research environments to support researchers' scholarly activities [17][18]. Several terms are used to refer to and distinguish among those services based on the main functionalities they provide, for instance, social bookmarking for researchers [19], online or social reference management (SRM) system [20], and academic social network.

---

* Part of this section is reprinted with permission from "The Evolution of Scholarly Digital Library Needs in an International Environment: Social Reference Management Systems and Qatar" by Hamed Alhoori, Carole Thompson, Richard Furuta, John Impagliazzo, Edward A. Fox, Mohammed Samaka, Somaya Al-Maadeed, 2013. Proceedings of the 15th International Conference on Asia-Pacific Digital Libraries, Copyright 2013 Springer International Publishing Switzerland. Part of this section is reprinted with permission from "Understanding the Dynamic Scholarly Research Needs and Behavior as Applied to Social Reference Management" by Hamed Alhoori and Richard Furuta, 2011. Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries, Copyright 2011 by Springer-Verlag GmbH Berlin Heidelberg.

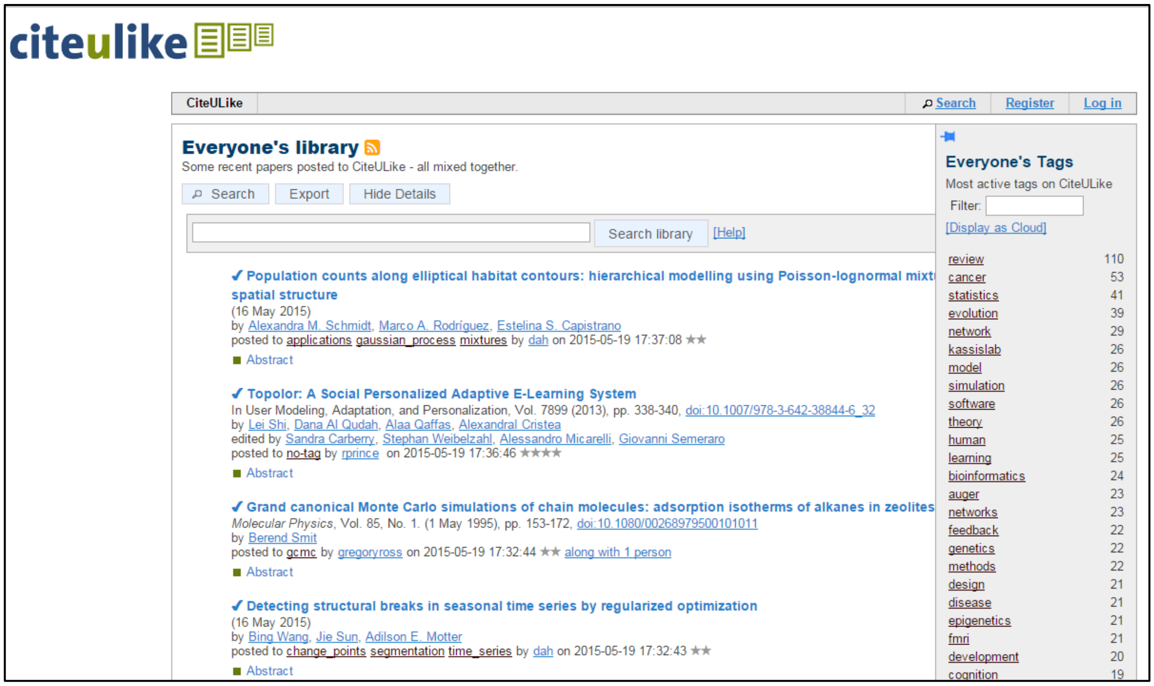


Figure 2.1 A screenshot of CiteULike

A number of popular SRMs and academic social networks have emerged and evolved, including CiteULike [21] (Figure 2.1), Zotero [22], BibSonomy [23], Mendeley [24], Academia.edu [25], and ResearchGate [26], used by millions of researchers worldwide. Such online services can serve as a reflection of scholarly big data.

As the number of scholarly products increases, and with the use of numerous social media tools during a research project's lifecycle, researchers' information needs, information-seeking behavior, and information use are not well-known or understood. The purpose of this section is to address this research gap and establish a better understanding of dynamic international scholarly information behavior by comparing the similarities among and differences between the behavior of researchers in the United

States (U.S.) and Qatar. Moreover, this study investigates whether academic social networks have any effect on scholarly information behavior.

2.2. Related Work

Numerous studies have been conducted in a range of disciplines in an effort to understand the scholarly information behavior of various groups. The disciplinary areas explored in this regard include architecture [27], astronomy [28][29], agricultural and biological sciences [8], business [30], chemistry [31][32], computer science [33], geoscience [34], humanities [35][36][37], law [38][39][40], mathematics [41], medicine and health sciences [42][43][44], public health [45], and veterinary medicine [46]. The groups include the Google generation [47], undergraduate students [48][49], graduate students [50][51], scientists [52][53], engineers [54][55][56], and academic scholars [57][58].

Several methods have been used to collect information about and to examine scholarly information behavior using quantitative studies [59][60][61] (e.g., surveys), qualitative studies [62][63] (e.g., interviews), ethnographic observational studies [64][65], and a combination of these. For example, Brown [66] used a combination of email survey and content analysis methods. Further, various studies used citation analysis to study researchers' information seeking behavior and information needs [67][68][69][70][71]. Other studies investigated usability evaluation methods [72], analyzed journals and article downloads [73], and used transactional log studies [74][75][76][77][78][79][80]. Overall, diverse models have been developed to capture and analyze information-seeking behavior [81][82][83].

A number of studies have shown that researchers are not aware of or familiar with some of the resources, services, and electronic search tools available to them through libraries and that researchers generally do not address their information needs with librarians [84][66][85][86]. To increase researchers' awareness, workshops and online tutorials [87][88] have been provided to support researchers' activities, such as the use of specific tools [89] (e.g., bibliographic management software).

Niu et al. [90] surveyed 2,063 academic researchers from several disciplines and research universities in the U.S. in an effort to better understand their information-seeking behavior. They found that differences in information-seeking behavior were clearer among disciplines and demographics than among universities. In a follow-up study, Niu and Hemminger [91] reported several factors affecting the information-seeking behavior of researchers, including demographics, psychological aspects, academic position, and discipline. Larivière et al. [92] found that doctoral students cite more recently published literature than faculty members.

Scholarly use of social media has been studied in blogs [93][94][95][96], wikis, and micro-blogging services, such as Twitter [97][98]. Recent studies have attempted to determine the influence of social media platforms on scientists and scholarly communities [99][100][101][102][103]. A few studies have investigated the effects of SRMs on scholarly communities [104][105][106]. In a study of the effects of social media tools on researchers at six universities in the United Kingdom, Tenopir et al. [107] found that around half of the 2,000 survey respondents read, viewed, and/or participated in at least one social media platform.

Gruzd and Goertzen [108] showed that the top reasons participants gave for using social media tools related to information-gathering activities. Among these reasons were to keep up-to-date on topics, to follow other researchers' work, to discover new ideas or publications, to promote current research, to make new research contacts, and to collaborate with other researchers. Mandavilli [109] found that a vital reason for using social media tools is to benefit from platforms that enable discussions of scholarly output to take place in a timely manner. Jeng et al. [110] studied a sample of users who had joined online research groups in Mendeley and found that they used the research features available more than the social features. Most of the studies conducted with the goal of learning about scholarly information behavior are either limited to a single university campus, language, culture, or tool or did not investigate the effects of using social media tools in academia [111][112].

2.3. Methodology

To build a thorough understanding of researchers' patterns, we conducted a mixed methods research study [113] whereby the qualitative research relied on interviews and the quantitative research on an online survey. Each interview lasted from 30 to 60 minutes. Both methods used the same set of questions. Before the interviews and the survey were administered, seven researchers reviewed the questions to assess the efficacy and completion time required. Minimal modifications were made based on their feedback. Participation in both studies was confidential and voluntary. The participants were made aware that they were free to withdraw at any time.

We investigated how changes in technologies available to research communities addressing social media use can benefit researchers, supporting their overall research progress and outcomes. Our central research questions were as follows:

- How do researchers select and use resources to search for scholarly content?
- How do researchers manage their scholarly content?
- How do researchers select collaborators, and what collaboration tools do they use?
- How do researchers stay up-to-date with new research relevant to their specialized area or to multidisciplinary areas?
- How do researchers measure the impact of research?
- Do social networking services have any influence on research communities?
- What are the current information needs of researchers?
- What difficulties do researchers encounter in the research process?
- What are the similarities among and the differences between the scholarly information needs and practices of researchers in the U.S. and those in Qatar?

In the U.S., eight randomly selected faculty members from different disciplines at Texas A&M University in College Station participated in personal interviews. Most of the interviewees supervised a research group with active researchers. The interviews started with a discussion of the current practices in the research group based on open-ended questions. Then, we moved to cover the unanswered questions from our list. For the survey, invitations were sent to participants in various university departments, and the resulting samples were random and independent.

In Qatar, the response rate for surveys was low, and given the absence of related studies conducted in Qatar, we focused on interviews that could provide more details. We used semi-structured interviews conducted in the interviewees' offices. The participants were mainly faculty members from Qatar University, which is the only national university in the country. We randomly selected a group of 32 faculty members engaged in research, of whom 21 participated in the study.

We refer to the first study as the *U.S. study* and to the second as the *Qatar study*. We refer to the U.S. study participants as PUX and the Qatar study participants as PQX, where X = {1, 2,…}. We used statistical hypothesis testing techniques. We mainly used the Pearson's chi-squared test (X^2) and analysis of variance (ANOVA).

2.4. Results

2.4.1. Survey

A total of 156 researchers participated in the online survey from the U.S. study, as shown in Figure 2.2. There were 124 male and 32 female respondents, and 64% were between 26 and 34 years old. The participating researchers represented 13 disciplines.

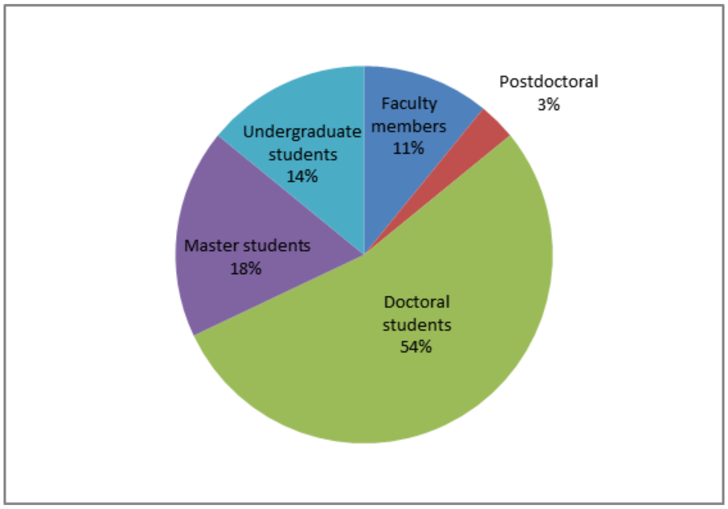


Figure 2.2 Distribution of participants in the survey

To archive information they discovered, the survey participants saved copies of articles and built personal article collections or repositories using a computer directory/folder, a reference manager, or an SRM. There was no significant relationship between the type of personal article collection and gender (Figure 2.3).

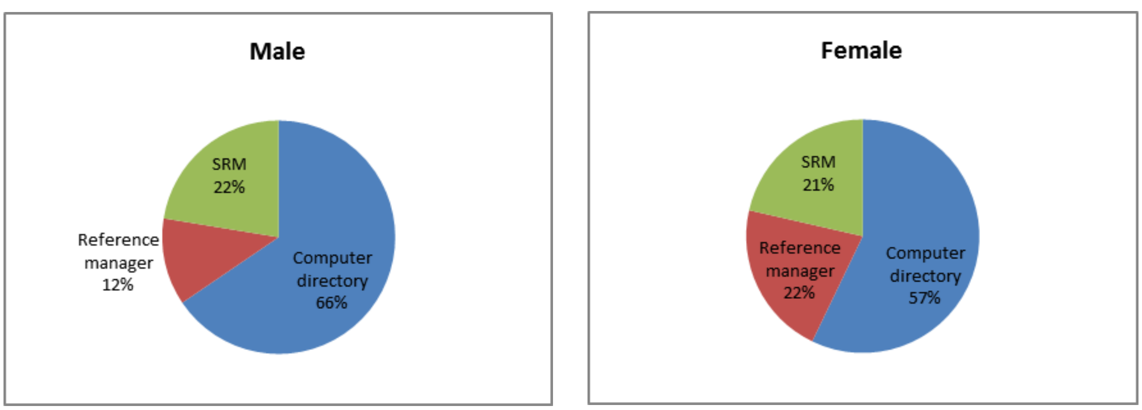


Figure 2.3 Type of personal article collection and gender

Figure 2.4 shows the type of personal article collection method employed and relevant academic status (e.g., student or faculty member). We found a significant relationship between these two factors ($p < 0.001$). A greater percentage of students than faculty members used SRMs to build personal article collections. This finding is in line with the findings reported in Mohammadi et al. [114] study in which PhD students were found to comprise the majority of Mendeley readers. The finding is also consistent with results reported in a study by Emanuel [115], which showed that graduate students use Mendeley (an SRM) more than faculty members do and that faculty members use EndNote (a reference manager) more than graduate students do.

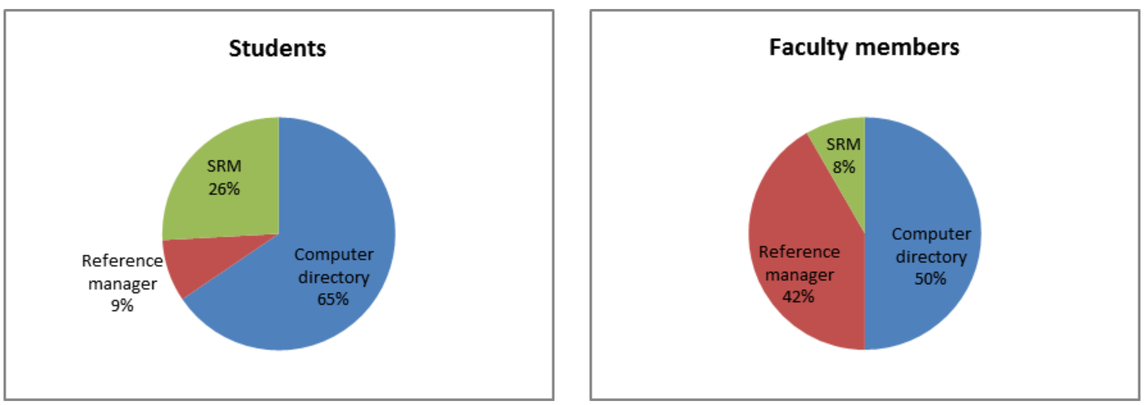


Figure 2.4 Type of personal article collection and academic status

Figure 2.5 shows nine disciplines and how researchers manage their scholarly article collections. We found a significant relationship between discipline and type of personal article collection ($p < 0.001$). The natural science participants used SRMs as their main approach to building a personal article collection, and none of them used computer directories to build a personal article collection. All the economics and mathematics researchers in the study built personal article collections using computer directories only.

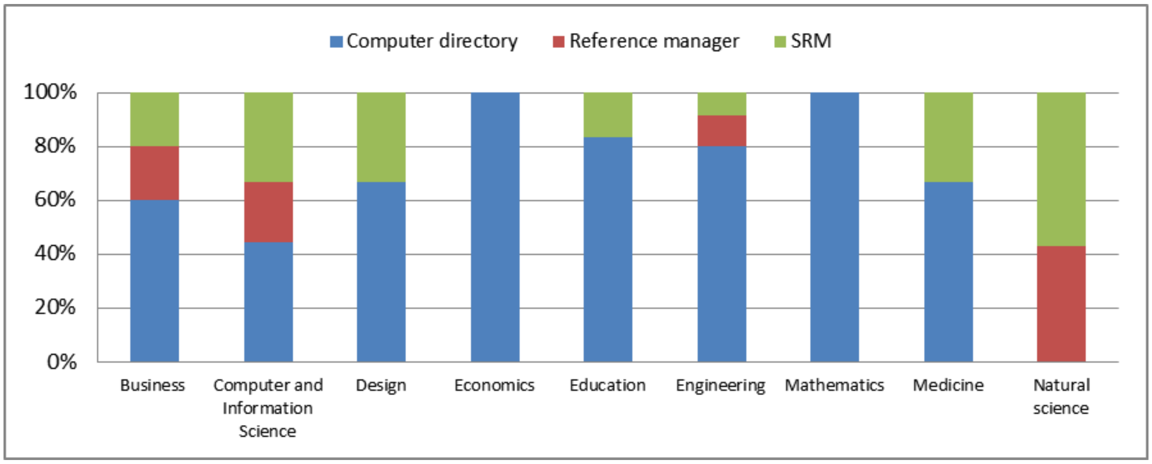


Figure 2.5 Comparison of using different personal article collection types in 9 fields

We considered the influence of the type of personal article collection on other scholarly activities. For example, we found that users of SRMs differ significantly from non-users of SRMs in regard to how they search for articles ($X^2 = 44.31$, df = 4, $p < 0.001$). Whereas most researchers used general or specific search engines, 40% of SRM users searched within SRMs. The participants explained that they use SRMs to search as such platforms have newer and more relevant results and allow them to connect with like-minded researchers. Similarly, Hallmark [116] showed that researchers in academia, government, and industry continue to develop new approaches to search for information in accordance with their needs.

Users of SRMs also used tags more often than other users. We found a significant relationship between SRM use and tag use ($X^2 = 19.032$, $df = 1$, $p < 0.001$). SRM users were able to find more articles related to their research interests than other users. However, there was no significant relationship between using SRMs and finding related topics.

Publication overload, which results when a researcher cannot keep abreast of the quantity of publications in his/her area of study, is a major challenge for most researchers (78%)—even for SRM users. However, there was no significant relationship between publication overload and type of personal article collection ($X^2 = 0.79$, $df = 2$, $p < 0.05$) or between publication overload and the ways in which survey participants organized their articles ($X^2 = 1.35$, $df = 1$, $p < 0.05$); i.e., whether they used directories, tags, and/or visual tools. Some SRM users showed an interest in using visual tools, but again, there was no strong evidence of a relationship using SRM and visual tools.

Survey participants who used directories noted they became disoriented more often when navigating between articles. Additionally, we found a significant relationship between the type of personal article collection and the tendency of the survey participants to become disoriented when reading and navigating between articles ($X^2 = 12.71$, $df = 6$, $p < 0.05$). We found another significant relationship between the type of personal article collection and writing notes on hard copies of articles ($X^2 = 5.64$, $df = 1$, $p <0.05$). Those who wrote notes on hard copies constituted 68% of those who used directories, 50% of those who used reference managers, and only 19% of those who used

SRMs. Furthermore, we found a significant relationship between the use of SRMs and making notes within SRMs ($X^2 = 17.03$, df = 1, $p < 0.001$).

We also found a significant relationship between the type of personal article collection and the first approach that researchers used to retrieve articles (i.e., searching or browsing) they had recently read ($X^2 = 9.98$, df = 2, $p < 0.05$). Those who retrieved articles by searching constituted only 31% of those who used directories, 50% of those who used reference managers, and 63% of those who used SRMs. There was a significant relationship between the type of personal article collection and whether or not the researchers collaborated with other researchers ($X^2 = 6.82$, df = 2, $p < 0.05$). Researchers who use reference managers and SRMs collaborated with more researchers than those who used directories.

Many researchers (67%) collaborated with others, for one or several of the following reasons: to share and expand knowledge, make new connections, to increase the possibility of securing funds, to become more motivated, to speed up the research process, or to publish more. The researchers who did not collaborate provided different reasons, including being busy with their research, finding it hard to compile or synchronize the work, or not knowing other researchers with similar interests.

Finally, we found strong evidence that the type of personal article collection had an effect on the satisfaction of researchers when searching for articles ($F = 37.80$, $p < 0.001$), retrieving articles ($F = 4.67$, $p < 0.05$), and organizing articles ($F = 4.66$, $p < 0.05$). A summary of the findings is presented in Table 2.1 ($p < 0.05$ = *, $p < 0.001$ = **, no significance = -).

Table 2.1 Summary of the relationships tested

Relationship tested in a scholarly activity	Significance
1) SRM users and	
a. searching for articles	**
b. using tags	**
c. finding related articles	-
2) Type of personal article collection and	
a. gender	-
b. academic status	**
c. discipline	**
d. publication overload	-
e. tendency to become disoriented	*
f. the writing of notes on hard copies of articles	*
g. first approach to retrieving articles	*
h. collaboration with other researchers	*
i. satisfaction with searching for articles	**
j. satisfaction with retrieving articles	*
k. satisfaction with organizing articles	*

2.4.2. Interviews

2.4.2.1. Searching for and reading scholarly content

In general, the interview participants described their reliance on well-known journals, conferences, bibliographic databases, and academic digital libraries to search for articles. A number of participants used Google Scholar, and some of these complained that this engine returned some articles that were unrelated to their search queries. In line with previous findings [117][118], the present study shows that researchers encountered some difficulties locating information of interest:

> *"I know the information is there, but I do not know how to reach it in a short period of time." (PU1)*

The participants differed in terms of their reading habits, but generally agreed that they skim the paper first by reading its abstract, conclusion, or results section before deciding whether to read the entire paper. Some reported that they became disoriented when navigating between different papers and references, whereas others, those who kept notes and focused on high-impact papers, did not report becoming disoriented. The participants generally agreed that they discontinue reviewing the literature when they have enough information for their purpose and/or when the content becomes repetitive. This finding is in accord with findings from studies of the information-seeking behavior of art administrators [119] and organizations [120].

Consistent with the Ellis model [121] and previous findings [122][123], chaining, i.e., following references from one article to another, was shown to be a common behavior and an important discovery method for researchers in the present study:

> *"During my reading of an article, I jump to skim the cited articles, and around 10% of the time, I would just neglect the initial article(s) after finding more interesting and related articles to my work." (PQ4)*

Most of the participants noted that they had come across at least a few articles later that would have added value to their completed or published work had they known the articles existed. Others complained that sometimes they were unable to locate articles they already knew of or had even read:

> *"I usually do not succeed in finding all related work, especially those that I skim, and I did not print nor read them." (PQ9)*

Several participants complained about redundant results during the search process:

> *"I would like to have a way to remove the previously viewed results from my new search results or when checking for new citations. Worse than that is when I get some search results that are already stored in my articles collection or reference manager and I start to view them again since my collection is huge and I cannot remember all articles." (PU2)*

2.4.2.2. Organizing and retrieving scholarly content

In organizing articles, some participants reported that they print out copies of articles. When asked why they did not move to electronic copies, they responded that they had been using this approach for a long time and did not want to jump from tool to tool:

> *"I print all the papers I need and organize them using authors' names. Although it may take some time to find what I need, but this way has worked for me since my graduate school." (PU3)*

A number of participants felt satisfied with organizing their papers and notes using computer folders and text files:

“I have been using folders to organize my papers and notes based on projects. I know all my folders, and when I need anything, I can go back to the project and to the subfolders.” (PU5)

One participant even used a general organizing tool:

“I am happy using my old file organizing tool version 1.0.” (PU6)

Several participants used reference managers and shared references among their groups. However, others, when asked why they did not use a reference manager tool, replied that they were concerned with the time needed to learn how to use the tool and the possibility of delaying their work:

“I have used a free reference manager provided by the university library. It was good, but it needs a license and continuous updates, which delay my work, especially when I move between several places.” (PU6)

Reference managers had become an integral assessment tool for several participants. For example, one offered the following rationale for using this kind of tool:

“I have around 12,000 articles, and I am daily adding a few more. I also share some with other scholars.” (PU4)

Some participants wrote notes on hard copies of articles or within their reference managers. Others preferred to use emails or online note-taking sites. A few even used text files and attached all saved articles, notes, or ideas to them. At least one researcher relied extensively on memory to locate a paper or a saved note:

“I have a strong memory, so I know most of my printed papers and the attached notes.” (PU1)

To keep up-to-date, some researchers noted that they repeat manual searches:

> *"I repeat some searches from time to time and check if there are any new articles to read." (PU5)*

2.4.2.3. Research collaboration and social platforms

All the faculty members collaborated on local or international levels, and several were engaged in multidisciplinary collaborations. Collaboration for them was usually performed through face-to-face meetings or by using communication tools (e.g., email), videoconferencing applications (e.g., Skype), and online file storage services (e.g., Dropbox):

> *"When conducting research in a multidisciplinary area, we are learning a new language and new skills. We try to learn what the other group is doing, and at a later point, each group will raise questions that neither group thought of before." (PQ8)*

Other participants were not satisfied with collaborating online:

> *"Even though we have regular online group meetings, we share files and results, but the collaboration is not moving as expected. Our research assistant is going to visit the other university this summer for a face-to-face collaboration." (PQ14)*

Furthermore, the participants collaborated with each other in order to expand their knowledge and expedite their work. Collaborators were selected for their expertise, reliability, and ability to work in a team. Some of the participants did not know how SRMs work, and they refused to spend time exploring them:

"I am busy with my work and getting my tenure. I do not want to spend time using an SRM and adding friends so that I can get article recommendations." (PU3)

A few researchers expressed regret about their lack of awareness regarding SRMs. However, SRM users expressed concerns about the accuracy of bibliographic data:

"I usually found some errors, missing bibliographic data or duplicate social bookmarks. So, I usually verify its data from the article's published press website." (PU8)

Most of the researchers were aware of or had used SRMs to some extent. One senior researcher took a position against using social networking services:

"All social media tools are distracting and produce noise, including the academic ones." (PQ16)

2.4.2.4. Publication overload

A number of the faculty members suffered from publication overload. Additionally, several complained that publication overload was having a negative effect on their research assistants:

"Although I spend enough time in explaining to the research assistants the research problem, some of them get distracted by publication overload and come back with nonrelated articles." (PU7)

"Some new research assistants are distracted by the huge amount of literature, and they spend a long time just to find out later that they were reading low-quality articles." (PQ10)

After learning that several research assistants had been distracted from their originally assigned research task, PQ12 found a temporary solution by creating a reading list for each new research assistant.

2.4.2.5. Scholarly impact

To gauge the importance of an article, researchers said they read and evaluate it. Citations were considered a secondary factor in determining the value of an article. When asked how scholarly impact should be measured, one participant suggested using the PageRank algorithm:

> *"The impact of an article should not be measured by summing up all citations, but by knowing the reputation of the researcher who cited the article." (PU8)*

Others were against using citations for evaluation purposes, such as one senior faculty member:

> *"The citations contain some politics in them more than science. Therefore, I think the real impact of research outcomes should be measured on how the research affected the community and human life rather than calculating a number." (PQ3)*

Although researchers sought work related to their interests from top journals, they did not consider citation-based journal rankings to be a primary measurement:

> *"I submitted a manuscript to a journal, and it was rejected, but I knew that the content and results were good. Therefore, I resubmitted it to another journal with a higher impact factor, and it was accepted." (PQ14)*

2.4.2.6. Specialized scholarly needs

The participants who used bibliographic management software sought a comprehensive solution with the ability to store all versions of articles, source codes, spreadsheets, presentations, posters, white papers, LaTeX files, Matlab files, and reports:

> *"I collect images of chemical formulas and store them inside documents. I also add notes near them for later retrieval." (PQ21)*

In terms of receiving recommendations for articles, some of the survey participants wished to receive recommendations more in line with their current research direction:

> *"Article recommender systems usually provide recommendations related to articles that I have added to my collection a few months or years prior, while I would like to get recommendations related to my current research interests." (PQ1)*

Researchers from both studies looked for advanced research tools capable of assisting them in collecting, summarizing, and analyzing the results from research articles. A number of participants from both studies avoided organizing their articles, even though they regularly failed to locate articles they had read previously. Several researchers mentioned that they would like to receive recommendations for scholarly venues and scientific events related to their work.

2.4.3. Further Discussion

We studied the scholarly practices of 25 faculty members working in the U.S. (8 through interviews and 17 through surveys) with 21 working in Qatar, as shown in Figure 22.6. We compared the scholarly activities of researchers who used SRMs and searched within them, built a personal article collection, took notes, collaborated with other researchers, used tags to organize articles, and were affected by publication overload.

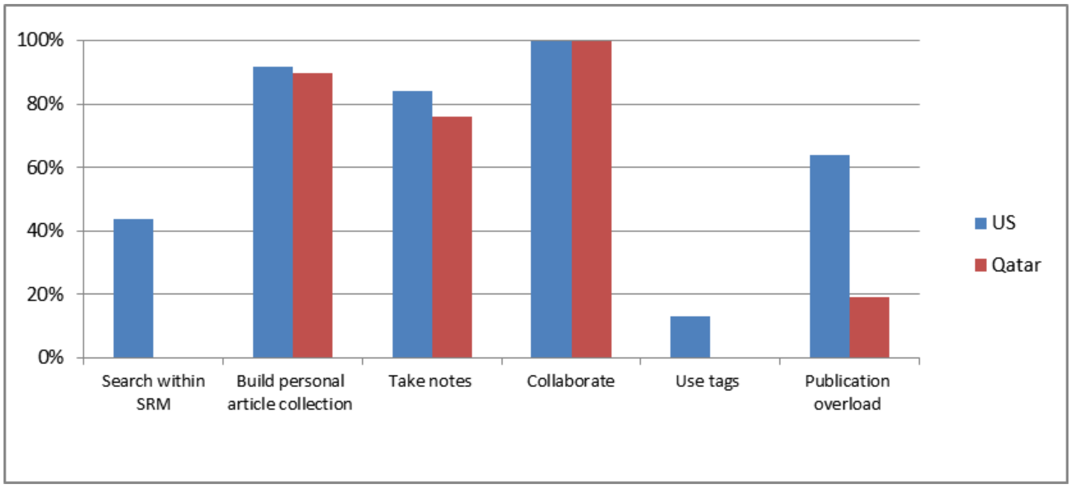


Figure 2.6 Comparison between the scholarly activities of faculty members working in the US and Qatar

In the U.S. study, we found a significant relationship between the use of SRMs and searching for articles. However, none of the participants in the Qatar study who used SRMs used them for the purpose of searching. None of the participants in the Qatar study used tags to organize their collections, whereas 13% of the U.S. study participants did use tags. Publication overload affected 64% of the faculty members in the U.S. study, whereas only 19% in the Qatar study noted being affected. One possible explanation is that most of the participants in the Qatar study focus on selected journals and conferences, whereas those in the U.S. follow several scholarly venues and multidisciplinary research areas. Similar to the U.S. study in which 88% of research assistants were affected by publication overload, several faculty members in the Qatar study noted that their research assistants were affected by publication overload.

We also found other similarities between the U.S. and Qatar studies. Unlike some previous studies that note differences between international students' information-seeking behavior [51], our findings show that in both studies some participants used similar scholarly resources, collaborated with other researchers, and used more than one method to build personal article collections and write notes.

The extent of the reluctance to use social media tools for scholarly purposes and to switch to new research assessment tools are also similar among faculty members in the two groups, which is consistent with results reported in other studies [124][125]. The reasons for this reluctance include learning curve time, concerns about delaying research, time needed to organize and update data, accuracy of bibliographic data, insufficient benefits, and high noise and distraction level.

Although more students than faculty members used the research assessment tools that support collaboration, not all the students collaborated during the research process, whereas all the faculty members collaborated. This finding indicates that students may not be using the available research tools effectively.

The survey results illustrated the vital effects of using SRMs, which differed from the interview results. One explanation for this difference could be that the majority of the survey participants were students and were more willing to experiment with new tools and technologies. Although the interviews helped us to provide clarifications for researchers and ask follow-up questions, some details may have remained hidden.

During the interviews, the researchers may have preferred not to mention the difficulties they experience because difficulties of this nature could be interpreted as a weakness, whereas the anonymous nature of the survey may led the participants to become more comfortable describing their difficulties and needs. This may help to explain the apparent differences between the two studies on some factors, such as publication overload.

3. MINING ALTMETRICS AT THE ARTICLE-LEVEL, JOURNAL-LEVEL, COUNTRY-LEVEL, AND ACCESS-LEVEL*

3.1. Introduction

Typical research dissemination methods include self-archiving preprint or post-print publications, presenting papers at conferences, publishing in NOA or OA journals, and sharing results with research groups. With recent and continuing research budget cuts, many research institutions have canceled costly subscriptions to journals [126]. Moreover, researchers may not be able to attend many related conferences or follow the vast range of publications available. Freed from subscription barriers, OA articles make knowledge more readily accessible to researchers and the general public, where the findings of the scholarly community become more visible.

By zooming out, we can see countries collaborate, compete, and compare their scientific production with other countries [127]. A country's reputation for research and development is important in terms of its relative ability to attract top scientists from around the world who can prepare young researchers, foster economic development,

* Part of this section is reprinted with permission from "On the Relationship between Open Access and Altmetrics" by Hamed Alhoori, Sagnik Ray Choudhury, Tarek Kanan, Edward Fox, Richard Furuta, C. Lee Giles, 2015. Proceedings of the iConference 2015, Copyright 2015 is held by the authors. Part of this section is reprinted with permission from "Do Altmetrics Follow the Crowd or Does the Crowd Follow Altmetrics?" by Hamed Alhoori and Richard Furuta. Proceedings of 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL), Copyright 2014 IEEE. Part of this section is reprinted with permission from "Altmetrics for Country-Level Research Assessment" by Hamed Alhoori, Richard Furuta, Myrna Tabet, Mohammed Samaka, Edward A. Fox, 2014. Proceedings of the 16th International Conference on Asia-Pacific Digital Libraries, Copyright 2014 by Springer International Publishing Switzerland.

open doors to international collaboration, create new jobs, and improve the quality of life for citizens and residents. Governments require that their dedicated GERD be utilized effectively and transformed into desirable outcomes [128] such as articles, patents, software, data, products, and services. However, despite this range of possible desirable outcomes, articles and citations have remained the dominant indicators of scholarly performance for researchers, journals, universities, and countries [129][130]. Citation analysis is a frequently used traditional approach to measuring research impact. However, citations may not exist for newly published articles or for articles that have local or limited regional benefit. Further, they have several limitations and cannot be used to measure the holistic impact of scholarly outcomes.

In order to maximize return on investment, policymakers, research funding agencies, and research communities are assessing various approaches to determine how public and private funds are being used, to measure the comprehensive impact of research, including both its scientific and social impacts, and to benefit from the research experiences of other nations [131]. Since January 2011, the National Science Foundation (NSF) has required grant proposals to include a data management plan, i.e., a "supplementary document that describes how the proposal will conform to NSF policy on the dissemination and sharing of research results" [132]. In December 2012, a group of editors and publishers of scholarly journals announced the San Francisco Declaration on Research Assessment [133], which recommends looking at a variety of metrics. Further, in January 2013, the NSF shifted its evaluation from a publication-based to product-based assessments [134].

In February 2013, the United States Office of Science and Technology Policy announced that it was expanding public access to the results of federally funded research [135]. In the UK, the higher education funding bodies have decided that "the impact element will include all kinds of social, economic and cultural benefits and impacts beyond academia, arising from excellent research" [136]. Furthermore, the grant-application forms for the UK Medical Research Council "specifically ask researchers how they intend to manage and share the results of their work, and to outline their productivity beyond published papers" [137]. And, in March 2014, the PLOS journals instituted a new data policy that requires authors to submit their data with their manuscripts, and in the case of publication, make the data publicly available [138]. Research evaluation is moving beyond traditional scholarly metrics and is increasingly taking into consideration the social, cultural, environmental, and economic impact of research [139][140][141].

Social media platforms enable researchers to distribute and discuss their results online, thereby widening the audience of readers who can study and measure the results and shortening the timeline for information to become available. An increasing amount of scholarly content is shared and discussed daily on social media platforms [142][143]. For example, it is estimated that the number of research articles shared on social media is increasing at the rate of 5–10% per month [144]. Social media platforms are playing an important role in the research lifecycle [101] inasmuch as researchers are using them for a number of purposes: to stay abreast of developments in their fields, to discover related work [108], to share and discuss research data and results [145], to connect with other researchers and citizen scientists, to collaborate online, and to obtain early feedback on their own work.

Online social interactions create traceable footprints and new data. By analyzing research use of social media platforms, researchers can identify users who are interested in their work and even determine the disciplines, universities, and/or countries with which those users are associated. As a result, these new models reveal previously unknown metrics and create new opportunities and challenges.

As research to date has focused on a narrow spectrum of social media platforms, little is known about the coverage, usage, distribution, validity, and trustworthiness of different platforms in research activities. Such information would have broad benefits for researchers interested in exploring online platforms to find a suitable environment for their scholarly activities; for bibliometricians interested in selecting platforms for measuring altmetrics; and for editors, publishers, libraries, research agencies, and social media platforms interested in providing better services to research-oriented communities. Further, few studies have examined the relationship between scholarly productivity and altmetrics at the country level [146]. Moreover, it is not clear whether altmetrics can be considered a universal measurement tool given that Internet access and the use of social media tools vary from one country or region to another.

In this section, we aim to answer the following research questions:

- How do social media platforms differ in terms of their coverage, usage, and distribution of scholarly works?
- How do altmetrics differ at the article, journal and country levels?

- How can we build and validate a comprehensive journal social-metric?
- Has the influence of journal rankings on researchers and general readers extended from scholarly communities to online communities?
- Can altmetrics support an assessment of the research of various countries?
- Do OA articles receive or generate higher altmetrics than NOA articles?
- Do NOA and OA articles published in the same journal and year receive different altmetrics counts?
- Is there a relationship between scholarly impact (citation count) and social impact (readership count [147]) for NOA and OA articles?

This section is structured as follows: We discuss related work in Subsection 3.2. In Subsection 3.3, we describe the data collection and detail our approaches. In Subsection 3.4, we present and discuss our results.

3.2. Related Work

3.2.1. Article, Journal, and Country Level Altmetrics

In order to evaluate the return on investment and assist with science policy, researchers have investigated several factors associated with measuring and comparing countries on the basis of scholarly outcomes. These factors include the number of publications, the number of citations, GERD, and gross domestic product (GDP) [148]. Moya-Anegón et al. [149] found a correlation (R^2 = 0.687) between the GDP of Latin American countries in 1995 and the number of indexed articles from those countries in 1996. They also found a high correlation between GERD and the number of articles (R^2 = 0.865). Tasli et al. [150] found that the number of articles in dermatology journals from 1999 to 2008 correlated with the GDP, population, and h-index of OECD countries. Meo et al. [151] found that GERD, number of universities, and number of scientific indexed journals correlated with the publications, citations, and h-index in various science and social science fields.

Research communities have long complained about the use of only one measure, such as citations, to determine the impact of scholarly entities [152], and alternative measures have been proposed. Neylon and Wu [153] found that various usage-based metrics, such as downloads, comments, and bookmarks, can be used to measure article and journal impact, and that each of these metrics has benefits and limitations. Bollen et al. [154] concluded that “the notion of scientific impact is a multi-dimensional construct that cannot be adequately measured by any single indicator.” At TPDL 2013, Borgman stated that “being cited in a tweet is a citation” [155].

Whereas citations measure an impact within scholarly boundaries, alternative metrics or altmetrics [156][157][158] provide the ability to measure a range of influences, including users who have read, shared, and/or discussed an article with others, but who have not formally cited it. Altmetrics was defined as “the creation and study of new metrics based on the social web for analyzing and informing scholarship” [159]. Considered complementary to traditional citation metrics, altmetrics are diverse in nature such that they can be used to measure the impact of a range of scholarly products, among which are articles, journals, books, datasets, software programs, and presentations. PLOS proposed Article-Level Metrics (ALM) [160], i.e., a comprehensive set of research impact indicators that include usage, citations, social bookmarking, dissemination activity, media, blog coverage, discussion activity, and ratings. The Usage Factor (UF) [161] “explores how online journal usage statistics might form the basis of a new measure of journal impact and quality.”

An increasing number of academic digital libraries and publishers such as Nature [162], Springer [163], BMJ [164], Cambridge Journals Online [165], and Scopus [166], are providing altmetrics on their websites. An editorial published in *Nature Chemistry* concluded that “despite its limitations, Twitter is useful for quickly disseminating information to an audience who has chosen to listen” [167].

Several researchers have studied the relationship between citation-based and social-based metrics and found low to moderate correlations, suggesting a complex relationship between altmetrics and scholarly impact. For example, Thelwall et al. [168] found an association between tweets and citations. Haustein et al. [169] found that 9.4% of PubMed articles were tweeted, but that a low correlation exists between citations and tweets. Similarly, Shuai et al. [170] reported a positive weak to moderate correlation between citations and Wikipedia mentions. Waltman and Costas [171] found a weak correlation between citations and F1000 recommendations, whereas Costas et al. [172] reported weak correlations between citations and altmetrics, and disciplinary differences using altmetrics. Several studies [173][174] have found a moderate correlation between citations and Mendeley readership for various disciplines and journals.

Using a sample of 20,000 publications from Web of Science with altmetrics from impactstory.org, Zahedi et al. [175] found that Mendeley’s coverage was the highest among all altmetric sources. Holmberg and Thelwall [176] analyzed tweets from selected researchers across ten disciplines and found some disciplinary differences in how researchers use Twitter, such as type of tweets, re-tweets, sharing of links, and conversations.

Few studies have examined the use of non-citation-based metrics as an early indicator of the scholarly impact of articles and journals. Brody et al. [177] found a significant correlation between the citations and downloads of articles in physics, mathematics, astrophysics, and condensed matter. They used the data of articles downloaded from within six months after publication as a predictive feature. In [14], we proposed a venue-ranking approach based on data from an online reference manager. The data selected was one year older than the matched data from traditional rankings. We compared the proposed social-based metrics with journal rankings and found statistically significant correlations. Most previous studies have attempted to understand altmetrics using only a few measures and focused on the article-level but not on the journal-level and country-level explored in the present study.

3.2.2. Open Access and Altmetrics

Several studies have investigated whether OA articles receive more citations than NOA articles (known as the OA citation advantage). Lawrence [178] found a citation advantage for conference articles in the field of computer science that are freely accessible online. Similar results have been reported in other fields, such as philosophy, political science, electrical engineering, electronic engineering, and mathematics [179], physics [180], agriculture [181], and civil engineering [182]. Hajjem et al. [183] used articles published over a 12-year period from 10 disciplines: administration, economics, education, business, psychology, health, political science, sociology, biology, and law. They found that OA articles had more citations and that the OA citation advantage ranged from 36% to 172%, according to discipline and year.

Norris et al. [184] found disciplinary differences in the citation advantage of OA articles in ecology, applied mathematics, sociology, and economics. Xia et al. [185] found that multiple OA availability correlates with citation count. McCabe and Snyder [186] found an OA citation advantage of 8% on average, with differences depending on content quality. Other reasons reported for high citation rates of OA articles include preprint availability, quality bias, and selection bias [187]. Eysenbach [188] controlled for various confounding variables and found that OA articles were likely to be cited twice as often as NOA articles in the first 4–10 months after publication. Gargouri et al. [189] reported that OA articles were not subject to a quality bias, finding a high OA citation advantage for both self-selected self-archiving and mandatory self-archiving articles.

A number of studies have explored the effects of social media on the dissemination of research. Shuai et al. [190] found that the number of tweets citing preprints on arXiv.org correlated with the number of downloads and early citations. Allen et al. [191] posted sixteen PLOS ONE articles on Facebook, Twitter, LinkedIn, and ResearchBlogging.org on either a random release date or a control date. They found that the dissemination of research through social media led to an increase in the number of views and downloads. Haustein et al. [192] found that the coverage and readership of articles published by sampled bibliometricians were higher on Mendeley than on

CiteULike. In other recent studies, we found that the altmetrics were related to traditional journal rankings [10] and countries' scholarly outcomes [11]. Shema et al. [193] found that articles cited on blogs received more citations. The focus of research to date has been limited to the citation rather than the altmetrics advantage of OA, and studies in this area have not drawn on a wide range of online metrics. The present study explores both of these factors.

3.3. Data and Methods

3.3.1. Journal-Level Altmetrics

We downloaded a dataset of 820 science journals from Journal Citation Reports (JCR) 2013 based on citation count. The data included abbreviated journal title, ISSN, impact factor (IF), five-year impact factor (5-IF), citation count, article count, immediacy index, cited half-life, Eigenfactor, and article influence score. We matched each abbreviated journal title with its full journal title. We then paired our data with the full set of SJR journal rankings using both ISSNs and the full journal names, as some of the ISSNs did not match. We obtained the SJR, h-index, total articles (three years), total citations (three years), and total references. Next, we matched this data against data from almetric.com [144], which collects article-level metrics, and we downloaded the article-level altmetrics for the previous year.

The altmetrics from altmetric.com comprise posts or mentions of research articles on CiteULike, Mendeley, F1000 reviews, blogs, Twitter, Facebook walls, mainstream news outlets, Google Plus, Pinterest, Reddit, and sites running Stack Exchange (Q&A). As some of the JCR journals did not match with the SJR rankings or altmetrics, our dataset was narrowed to 785 journal titles, with 373,427 articles, which resulted in an altmetrics count of 13,221,827. We define a new social-based metric, *Journal Social Impact* (JSI), which represents the average number of posts or mentions of research articles on online platforms for a journal (j), as shown in Equation 3.1:

$$JSI(j) = \frac{\sum_{s \in S} \sum_{a \in A} a_s}{|A|} \tag{3.1}$$

s represents one of the social media platforms from the set S. a represents an article from the set of all articles A in a journal. $|A|$ denotes the total number of articles from a journal that were posted on online platforms. a_s represents the number of times an article a was posted in s by different users. We used Spearman's rank correlation coefficient, ρ(rho), to compare JSI and altmetrics with various citation-based metrics. We compared our collection of altmetrics with the altmetric.com score, which is a weighted score based on the volume, sources, and authors of online mentions. We also compared our collection of altmetrics with the number of social media platforms on which an article had been mentioned.

3.3.2. Country-Level Altmetrics

We selected 35 developed and developing countries that published 2,000 or fewer indexed articles per year from January 1, 2010, to June 5, 2014. We included articles co-authored by researchers from different countries. We used Scopus to download the bibliometric data, including DOI, citation, and year of publication. We used only articles with DOIs, resulting in a total of 76,517 bibliometric records.

We obtained the h-index for each country included in this study from SCIMago [194]. We matched Scopus DOIs with data from altmetric.com for each article. We then compared citation-based data with five types of altmetrics data sources: Twitter, Facebook, mainstream news outlets, blogs, and Google Plus.

We downloaded the GDP, GDP per capita, number of Internet users, number of mobile users, and number of researchers per country from the World Bank's DataBank [195] for the years 2011 and 2012, as articles published in 2012 were likely to have been funded in 2011 or earlier. For the few countries for which GDP was not documented at the World Bank, we used data from the United Nations National Accounts Main Aggregates Database [196]. We used the latest GERD available for 2011 for each country from the World Bank. Similarly, the GERD for some countries was not documented at the World Bank. Therefore, for these countries, we used data from R&D Magazine [197].

Finally, we obtained data on the usage of social networks for countries from the World Economic Forum's Global Information Technology Report [198]. We used Spearman's rank correlation coefficient, ρ(rho), to compare countries on the basis of different metrics.

3.3.3. Access-Level Altmetrics

We randomly selected 23 NOA and hybrid OA journals from the top 100 journals from all fields as ranked by the 2014 Google Scholar h-index [199]. We used Scopus to download bibliographic information for 42,582 articles published in the selected journals between 2010 and 2014. From the downloaded articles, we selected only those that had

DOIs. We then used Google Scholar to determine which of our articles were OA or NOA, because Google Scholar was found to retrieve a higher percentage of OA articles than OAIster and OpenDOAR [200].

We modified a parser for Google Scholar [201] to read our collection of articles, conduct an article title search, and if available retrieve a direct link to the full text of each article (i.e., the search result link adjacent to the article title on the Google Scholar results page). We ran the parser on a computer that did not have access to any journal subscriptions. In general, for each article, the parser returned one of two results: a web link to the article (e.g., *.html* or *.pdf*) or no link at all. For seven of the journals, we found that Google Scholar returned many links to NOA articles; therefore, we excluded those journals. We removed duplicate articles as well as those retrieved by Google Scholar for years outside the 2010–2014 range, thus reducing the number of journals to 16 and the number of articles to 27,011.

We defined OA articles as those for which the search returned a link, whereas articles for which a link was not returned were flagged as NOA, as shown in Figure 33.1. Using a random sample of 400 articles, we tested whether the title of the returned article link by Google Scholar matched our query title and found an accuracy rate of 99.2%. Using other random samples of 400 NOA articles and 400 OA articles, we checked the accuracy of our classifications of articles as NOA or OA and found accuracy rates of 97.5% and 96%, respectively.

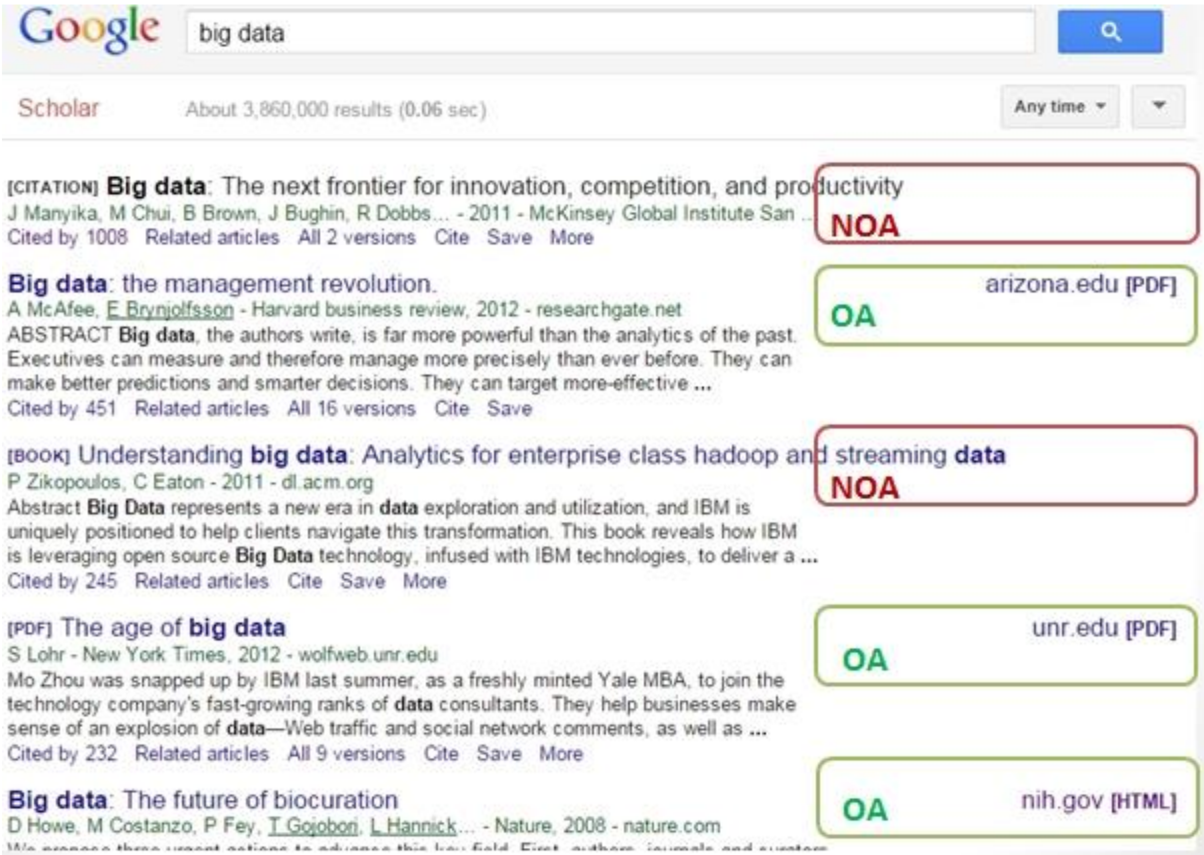


Figure 3.1 Classifying NOA and OA Articles using Google Scholar

We downloaded each journal's altmetrics from altmetric.com, which comprise mentions of articles on Twitter, Facebook, CiteULike, Mendeley, F1000, blogs, mainstream news outlets, Google Plus, Pinterest, Reddit, Sina Weibo, the peer review sites PubPeer and Publons, policy documents, and sites running Stack Exchange (Q&A). We then matched the articles using DOIs. We removed three sources of altmetrics—Pinterest, Q&A sites, and policy documents—due to insufficient data.

We defined the *OA Altmetric Advantage* (OAAA) for all types of altmetrics as shown in Equation 3.2. $\overline{OA}$ represents either the average number of articles that received an altmetric (article-based) or the average altmetric across articles (altmetric-based) for OA articles, and $\overline{NOA}$ represents the same for NOA articles.[1]

$$OA\ Altmetric\ Advantage(OAAA) = \frac{\overline{OA} - \overline{NOA}}{\overline{NOA}} \tag{3.2}$$

We compared NOA articles with OA articles based on altmetric type. We then compared articles with similar altmetric types that were published in the same year. In order to reduce the effects of platform, time, journal ranking (e.g., Impact Factor), and discipline, we extended the comparison by checking articles based on the altmetric type per journal per published year. We used the Mann-Whitney U test to check for significant differences between NOA and OA articles in regard to the altmetrics advantage. We used Spearman's rank correlation coefficient, ρ(rho), to compare the citation count with Mendeley readership. We used Mendeley because we found that it has a high usage and coverage of scholarly activities [10].

[1] For example, for a total of 1,000 articles, 400 OA and 600 NOA, and among them 40 OA and 30 NOA with a specific type of altmetrics (e.g., tweet), totaling 800 tweets for OA and 900 tweets for NOA. An article-based approach yields $\overline{OA}$ = 40/400 = 0.1, $\overline{NOA}$ = 30/600 = 0.05, and *OAAA* = (0.1 - 0.05)/0.05*100 = 100%. An altmetric-based approach yields $\overline{OA}$ = 800/400=2, $\overline{NOA}$ = 900/600=1.5, and *OAAA* = (2-1.5)/1.5*100 = 33.3%.

3.4. Results and Discussion

3.4.1. Coverage, Usage, and Distribution

As shown in Figure 3.2, Mendeley and Twitter have the highest coverage of articles shared on online platforms, whereas only 10% of the shared articles are covered in the mainstream news. Next, we found that Mendeley was the predominant platform on which research articles were shared, with 74% of the total altmetrics count, followed at some considerable distance by Twitter with 19%. The remaining 7% was distributed among all the other tested sites, as shown in Figure 3.33 Pinterest and the Q&A sites have the lowest levels of coverage and usage. Figure 3.4 shows that around 46% of our collection of articles was shared on two platforms.

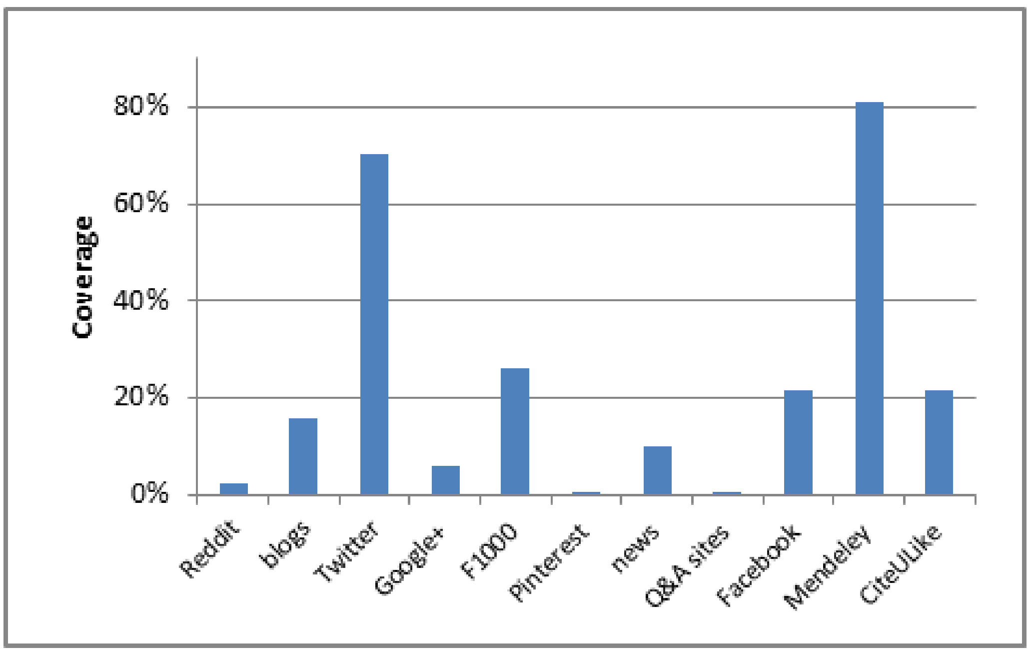


Figure 3.2 Coverage of research articles on various platforms

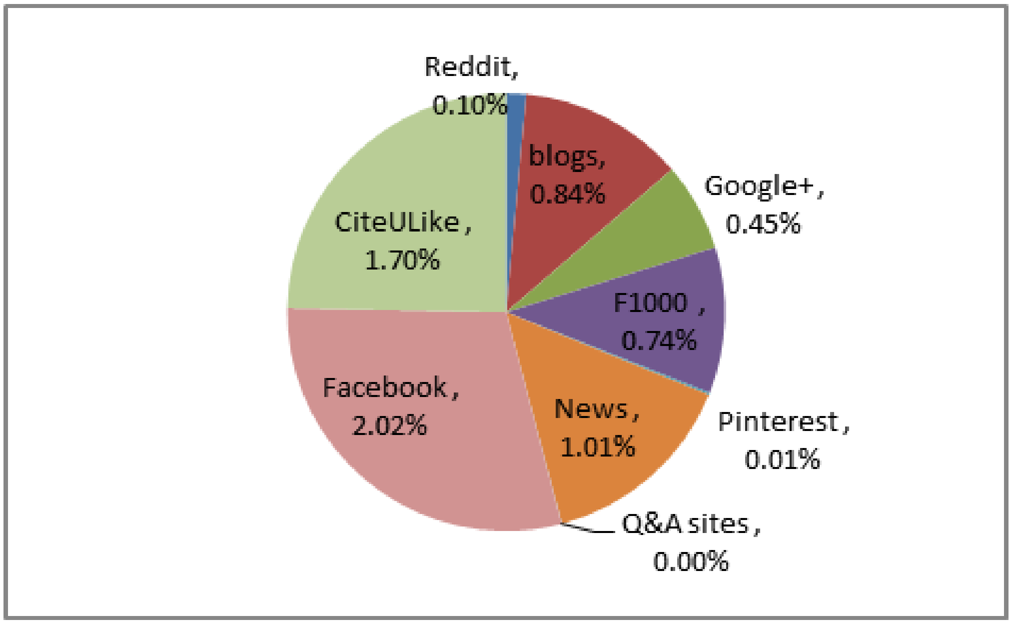


Figure 3.3 Research use of 9 online platforms (Distribution of altmetrics count across platforms)

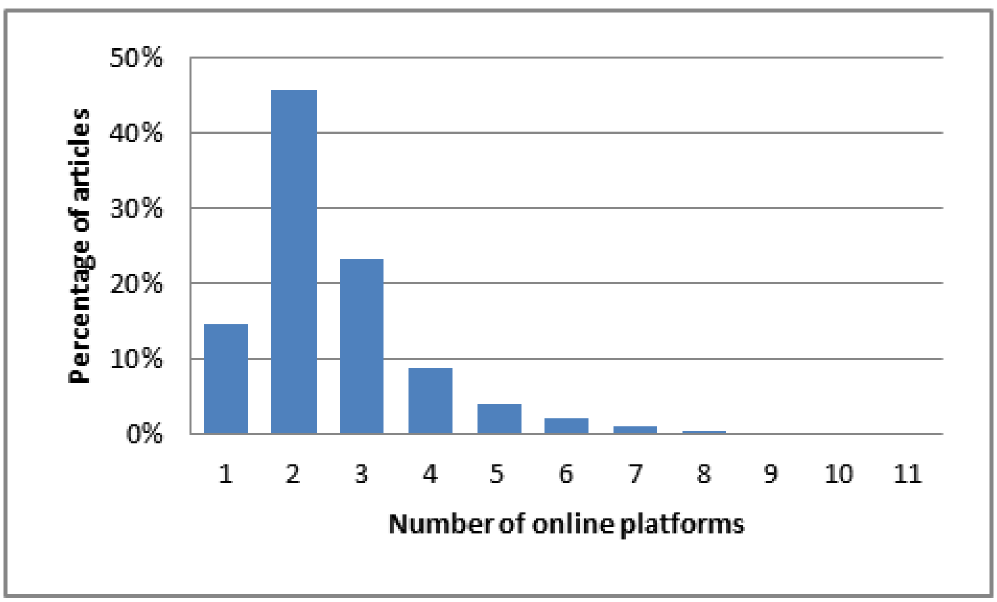


Figure 3.4 Distribution of articles across various platforms

3.4.2. Article-Level Altmetrics

We found that altmetrics at the article-level have weak correlations with citation-based metrics as shown in Table 3.1. The highest correlation was between Mendeley and the article influence score (ρ = 0.353, p < 0.01).

Table 3.1 Correlations between article-level altmetrics and traditional metrics.

	Citation Count	Impact Factor	5 -Year Impact Factor	Immediacy Index	Article Count	Eigenfactor Score	Article Influence Score	SJR	H-index
Altmetric Score	0.10	0.25	0.26	0.23	-0.05	0.15	0.27	0.22	0.13
Reddit	0.10	0.08	0.09	0.07	0.05	0.12	0.09	0.07	0.06
Blogs	0.13	0.20	0.20	0.19	0.00	0.16	0.22	0.19	0.13
Twitter	0.02	0.09	0.10	0.06	0.00	0.06	0.11	0.04	-0.03
Google+	0.09	0.11	0.12	0.10	0.03	0.12	0.13	0.10	0.08
F1000	0.11	0.26	0.26	0.26	-0.12	0.13	0.27	0.30	0.27
Pinterest	0.05	0.06	0.06	0.05	0.01	0.06	0.06	0.05	0.05
News	0.11	0.15	0.16	0.15	0.02	0.13	0.16	0.14	0.10
Q&A	0.03	0.04	0.04	0.04	0.00	0.03	0.04	0.03	0.03
Facebook	0.08	0.11	0.12	0.09	0.01	0.09	0.12	0.06	0.04
Mendeley	0.06	0.32	0.34	0.27	-0.15	0.14	0.35	0.35	0.19
CiteULike	0.10	0.26	0.27	0.24	-0.07	0.16	0.29	0.28	0.19

In general, the article-level altmetrics have weak correlations among themselves, except in a few cases, as shown in Table 33.2. In other words, articles that receive social attention on one online platform do not necessarily receive similar attention on other platforms. All correlations were significant at (ρ < 0.01). These findings show that article-level altmetrics measure a social impact that differs from scholarly impact.

Blogs have a weak correlation with news (ρ = 0.313). Further, Twitter showed a weak correlation with Facebook wall posts (ρ = 0.304), and Mendeley has a moderate correlation with CiteULike (ρ = 0454). F1000 showed a positive moderate correlation with Mendeley readership (ρ = 0.454) and a negative moderate correlation with tweets (ρ = -0.464), which shows the scholarly nature of online reference managers' data. The altmetric.com score has moderate correlations with blogs, tweets, news, and the number of platforms on which an article was mentioned (ρ = 0.570, 0.580, 0.488, and 0.526, respectively). The latter has moderate correlations with blogs, Facebook posts, Mendeley, and CiteULike (ρ = 0.469, 0.463, 0.585, and 0.577, respectively).

Table 3.2 Correlation matrix between article-level altmetrics

	Reddit	**Blogs**	**Twitter**	**Google+**	**F1000**	**Pinterest**	**News**	**Q&A**	**Facebook**	**Mendeley**	**CiteULike**	**No. Platforms (S)**	**Altmetric Score**
Reddit	1.00	0.21	0.18	0.26	-0.05	0.14	0.20	0.03	0.21	0.05	0.07	0.22	0.19
Blogs	0.21	1.00	0.16	0.24	-0.04	0.08	0.31	0.04	0.21	0.23	0.24	0.47	0.57
Twitter	0.18	0.16	1.00	0.26	-0.46	0.07	0.21	0.02	0.30	-0.15	0.01	0.28	0.58
Google+	0.26	0.24	0.26	1.00	-0.08	0.10	0.25	0.03	0.26	0.06	0.10	0.32	0.30
F1000	-0.05	-0.04	-0.46	-0.08	1.00	-0.01	-0.09	0.00	-0.17	0.45	0.21	0.20	-0.09
Pinterest	0.14	0.08	0.07	0.10	-0.01	1.00	0.06	0.03	0.09	0.03	0.03	0.08	0.07
news	0.20	0.31	0.21	0.25	-0.09	0.06	1.00	0.02	0.22	0.06	0.08	0.35	0.49
Q&A	0.03	0.04	0.02	0.03	0.00	0.03	0.02	1.00	0.02	0.02	0.03	0.03	0.03
Facebook	0.21	0.21	0.30	0.26	-0.17	0.09	0.22	0.02	1.00	0.05	0.06	0.46	0.27
Mendeley	0.05	0.23	-0.15	0.06	0.45	0.03	0.06	0.02	0.05	1.00	0.45	0.59	0.15
CiteULike	0.07	0.24	0.01	0.10	0.21	0.03	0.08	0.03	0.06	0.45	1.00	0.58	0.17
No. Platforms (S)	0.22	0.47	0.28	0.32	0.20	0.08	0.35	0.03	0.46	0.59	0.58	1.00	0.53
Altmetric Score	0.19	0.57	0.58	0.30	-0.09	0.07	0.49	0.03	0.27	0.15	0.17	0.53	1.00

3.4.3. Journal-Level Altmetrics

Some of the metrics we studied did not correlate with any of the others, such as cited half-life and total references, so we removed them from the results. Table 3.3 shows that most our collection of journal-level altmetrics have moderate correlations with journal citation count, h-index, and Eigenfactor, and weak correlations with other citation-based metrics. However, JSI has significant positive moderate correlations with IF, 5-IF, Immediacy Index, SJR, and article influence score. In addition, JSI has a higher correlation with 5-IF and article influence score than with the Immediacy Index, which shows that JSI has a stronger relationship with reputable journals that have a history of scholarly impact.

Among the journal-level altmetrics, Mendeley and CiteULike readers have the highest correlations with all the journal rankings, which shows that these online reference managers are more related to scholarly impact than other metrics. Mainstream news has the highest correlation with citation count and Eigenfactor, which indicates that research disseminated to the public by new providers is related to popular and quality journals. Again, all correlations measured here were significant at ($p < 0.01$).

Table 3.3 Correlations between journal-level altmetrics and traditional metrics

	Citation Count	IF	5-IF	Immediacy Index	Article count	Eigenfactor	Article Influence Score	SJR	H-index
Reddit	0.36	0.19	0.20	0.30	0.19	0.45	0.23	0.17	0.37
Blogs	0.48	0.23	0.25	0.36	0.25	0.57	0.30	0.27	0.51
Twitter	0.45	0.19	0.21	0.34	0.27	0.55	0.23	0.20	0.48
Google+	0.44	0.21	0.23	0.35	0.25	0.54	0.26	0.21	0.47
F1000	0.40	0.17	0.19	0.30	0.16	0.46	0.23	0.21	0.51
Pinterest	0.11	0.25	0.26	0.25	-0.01	0.21	0.26	0.23	0.18
News	0.51	0.14	0.16	0.30	0.37	0.59	0.17	0.17	0.45
Q&A	0.17	0.24	0.24	0.24	-0.05	0.20	0.24	0.23	0.23
Facebook	0.46	0.16	0.17	0.31	0.29	0.53	0.17	0.13	0.46
Mendeley	0.46	0.35	0.39	0.41	0.14	0.55	0.43	0.41	0.59
CiteULike	0.41	0.33	0.36	0.42	0.10	0.52	0.45	0.41	0.56
JSI	-0.04	0.58	0.63	0.46	-0.39	0.07	0.67	0.58	0.23

With the exceptions of Pinterest and the Q&A site, we found moderate to strong correlations between journal-level altmetrics, as shown in Table 3.4, which differ from the article-level altmetrics, for which we found weak correlations. The lowest correlations were between F1000 and Reddit ($\rho = 0.587$) and between F1000 and Google Plus ($\rho = 0.610$). The highest correlations were between Twitter and Facebook ($\rho = 0.914$) and between Mendeley and CiteULike ($\rho = 0.912$). Comparing article-level altmetrics from different disciplines seems like comparing apples to oranges, but comparing clustered altmetrics based on journals is like comparing apples to apples. General and academic social media platforms cluster together and present high correlations among themselves. All correlations were significant at ($p < 0.01$).

Table 3.4 Correlation matrix between altmetrics for journals

	Reddit	Blogs	Twitter	Google+	F1000	News	Facebook	Mendeley	CiteULike
Reddit	1.00	0.77	0.77	0.76	0.59	0.68	0.78	0.69	0.66
Blogs	0.77	1.00	0.87	0.82	0.68	0.82	0.84	0.85	0.83
Twitter	0.77	0.87	1.00	0.83	0.73	0.81	0.91	0.84	0.80
Google+	0.76	0.82	0.83	1.00	0.61	0.74	0.81	0.77	0.76
F1000	0.59	0.68	0.73	0.61	1.00	0.66	0.72	0.77	0.72
News	0.68	0.82	0.81	0.74	0.66	1.00	0.81	0.70	0.64
Facebook	0.78	0.84	0.91	0.81	0.72	0.81	1.00	0.77	0.72
Mendeley	0.69	0.85	0.84	0.77	0.77	0.70	0.77	1.00	0.91
CiteULike	0.66	0.83	0.80	0.76	0.72	0.64	0.72	0.91	1.00

The absence of high correlations between altmetrics and citation-based metrics shows the existence of differences between the scholarly sphere and the more general social sphere in regard to the level of importance attached to research. In addition, it can be explained that social attention measures new findings, public interest, gaming to the altmetrics system, or even spam targeting specific communities, such as the scholarly world [15].

3.4.4. Country-Level Altmetrics

The total number of articles cited (citations coverage) was significantly higher than the number of articles that received altmetrics (altmetrics coverage). However, by considering individual years, we found that altmetrics are increasing significantly, as shown in Figure 33.5. Moreover, articles published in 2014 received significantly more altmetrics (27%) than citations (10%). Among these articles, 22% have only altmetrics and 6% have only citations, which shows that altmetrics can work as an early social impact indicator. Fifteen percent of the articles were referenced via Twitter, 4% via Facebook, 2% via blogs, 1% via Google Plus, and 1% reached the mainstream news.

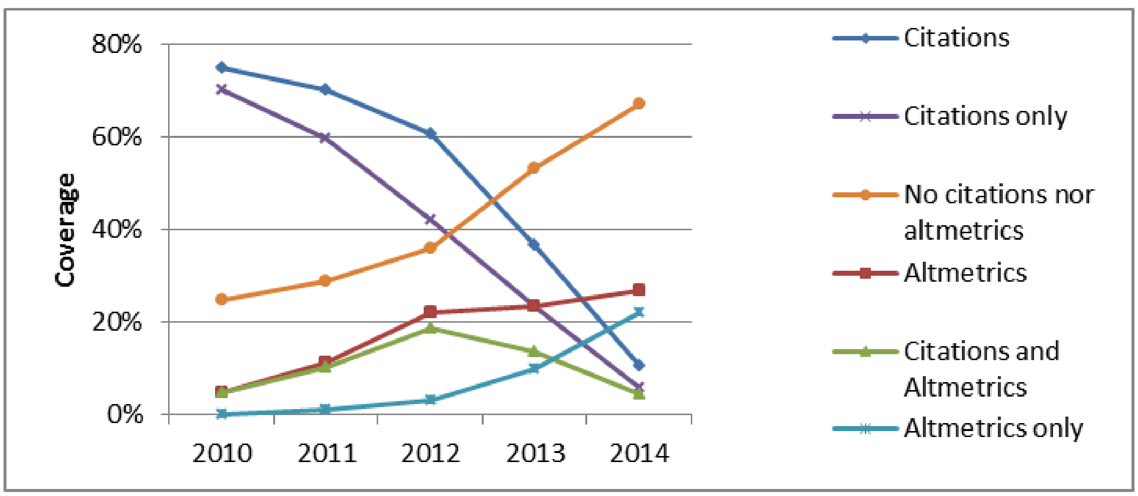


Figure 3.5 Coverage of citations and altmetrics from January 2010 to June 2014

Less than 20% of the articles received citations and altmetrics each year, which creates a challenge in regard to evaluating or validating impact using both metrics. Moreover, a very large proportion of the published articles did not have any citations or altmetrics, even a few years after publication. For example, in 2010, 25% of the articles had neither citations nor altmetrics, and for 2013 the figure was 53%. We found that the metrics for 2012 had similar correlations to the metrics for 2011. Therefore, we decided to report correlations based only on metrics from 2011, as shown in Table 3.5.

Table 3.5 Correlations between country-level altmetrics and traditional metrics

	GERD	Total articles	Total citations	H-index	Citations coverage	Altmetrics coverage	Internet users
GERD	1.00	0.75	0.67	0.63	0.72	0.61	0.47
Total articles	0.75	1.00	0.91	0.70	0.98	0.84	0.49
Total citations	0.67	0.91	1.00	0.79	0.95	0.94	0.42
H-index	0.63	0.70	0.79	1.00	0.75	0.83	0.33
Citations coverage	0.72	0.98	0.95	0.75	1.00	0.89	0.49
Altmetrics coverage	0.61	0.84	0.94	0.83	0.89	1.00	0.44
Internet users	0.47	0.49	0.42	0.33	0.49	0.44	1.00

The GERD had higher correlations than the GDP. The GDP per capita and citations per article had low correlations with other metrics; however, the h-index had strong correlations. The number of Internet users, the number of mobile users, and usage of social networks had low to moderate correlations, showing that altmetrics are not strongly related to the number of general users.

Individual altmetrics counts (e.g., scholarly tweets counts) and altmetrics coverage were strongly correlated with citations and citations coverage. The number of researchers was not available for ten of the countries. However, a comparison of the 25 countries for which data was available showed low correlations between the number of researchers and the other metrics. All correlations were significant at ($\rho < 0.05$).

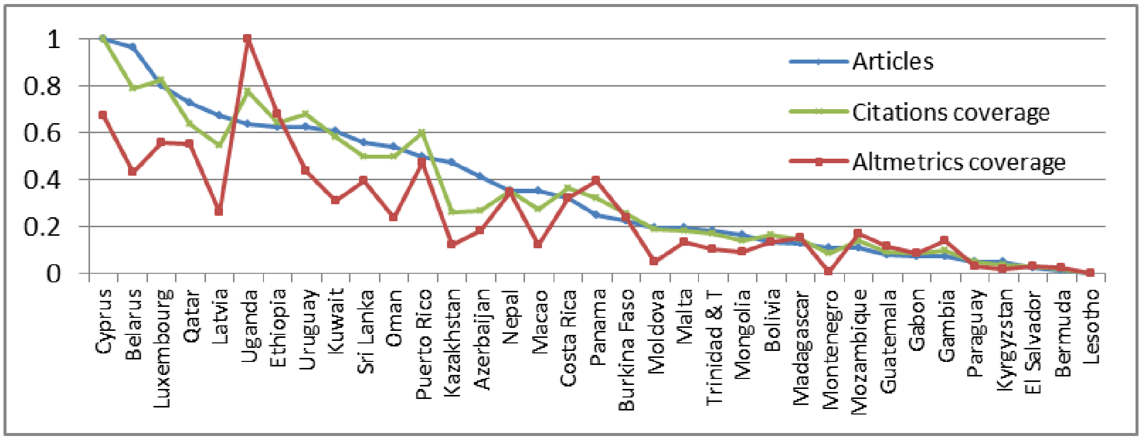


Figure 3.6 Countries' scholarly production impact and social impact based on normalized data

Figure 33.6 shows a high level of significant correlation ($\rho = 0.92$) between citations coverage and altmetrics coverage based on normalized data for all articles and years, which can help in predicting and validating both scholarly and social impact.

3.4.5. Access-Level Altmetrics

Of the 27,011 articles, 6,934 were NOA and 20,077 were OA. Figure 33.7 provides descriptive statistics of the articles that received various types of altmetrics, with their count, percentage, and access type. The vertical axis shows the percentage of NOA (gray columns) and of OA (light-blue columns) articles.

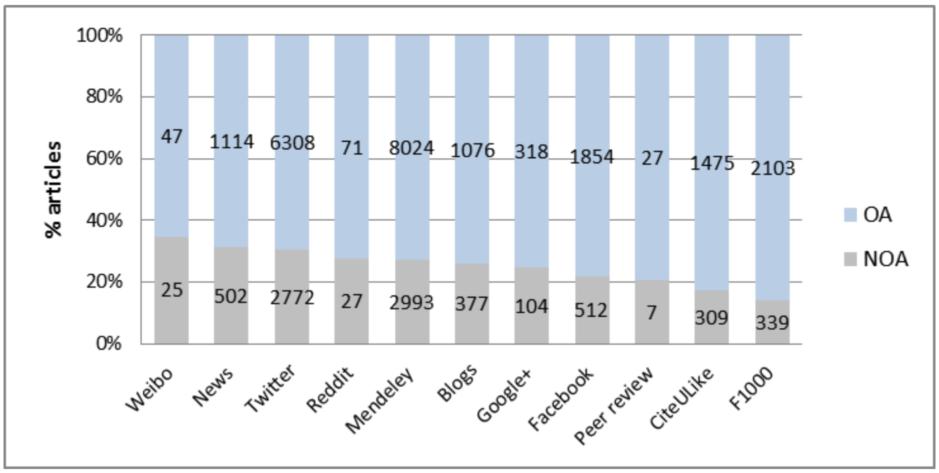


Figure 3.7 Distribution of NOA and OA articles across online platforms

We compared the NOA and OA articles that received altmetrics with those that did not, using an article-based approach (Figure 3.8) and an altmetric-based approach (Figure 3.9). Figure 3.8 shows the percentages of $\overline{OA}$ and $\overline{NOA}$ based on article count, and the right side shows an article-based OAAA, which is represented by the red curve. Six platforms did not show any article-based OAAA. However, a clear article-based OAAA is shown for both F1000 and CiteULike.

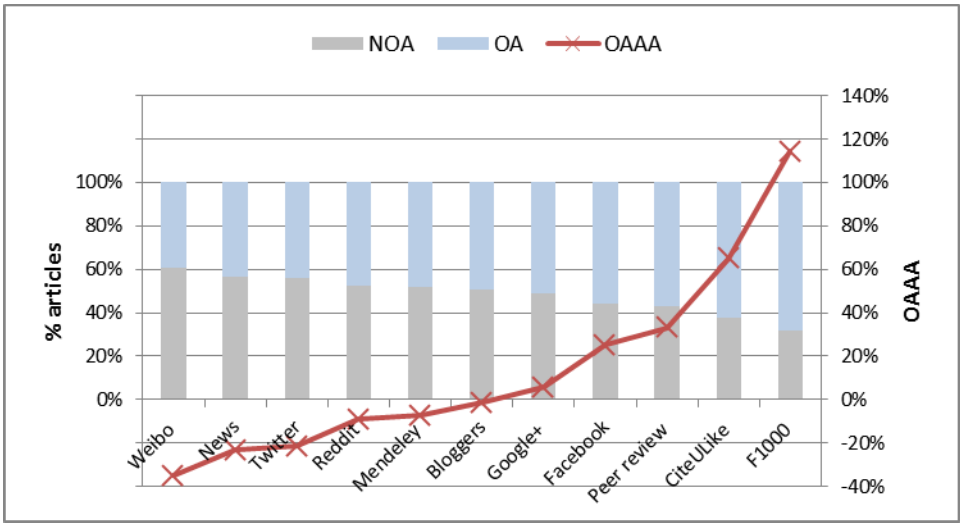


Figure 3.8 Article-based OAAA

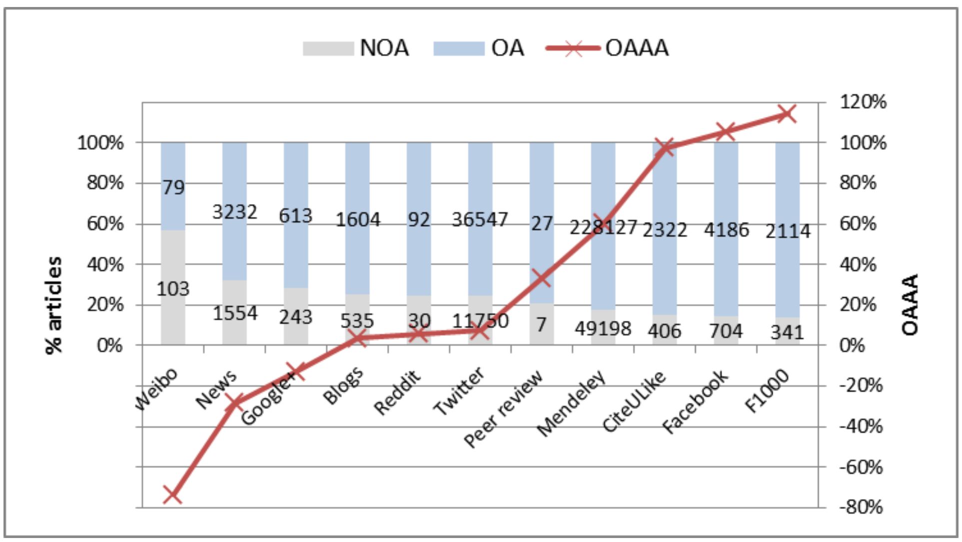


Figure 3.9 Altmetric-based OAAA

Figure 3.9 illustrates the distribution of altmetrics for NOA and OA articles. It shows altmetric-based OAAA on eight platforms, four of which are above 50%. Figure 3.8 shows that a higher percentage of OA articles received more altmetrics than NOA articles on F1000, CiteULike, Facebook, and peer review sites. Mendeley covers a slightly higher percentage of NOA articles (Figure 3.8), but the OA articles have 60% more readers (Figure 3.9). Academic social networks (e.g., F1000, CiteULike, and Mendeley) received high altmetric-based OAAA, whereas there was a clear difference between the general social media sites in terms of altmetrics received by NOA and OA articles. For example, Facebook covered a high percentage of OA articles and showed a high OAAA (105%). However, Twitter covered a high percentage of NOA articles, but OA articles received more tweets (7%), which might be the effect of publishers sharing NOA articles on Twitter more often than on Facebook. Google Plus, mainstream news outlets, and Weibo did not receive altmetric-based OAAA, which could be due to the effect of high impact articles published in high-ranked NOA journals [10].

Table 33.6 reports significant differences (p-value <0.05) between NOA and OA articles in terms of type of altmetrics and year. CiteULike and F1000 each showed a significant difference between NOA and OA articles for the years 2010–2013. However, no significant difference was found between NOA and OA articles for CiteULike or for F1000 in 2014, which could be due either to insufficient data, declining OA advantage [202], or reduce usage of such sites in scholarly dissemination and a possible shift to other sites. Twitter and Mendeley showed significant differences between NOA and OA articles in all the years studied, with the exceptions of 2011 for Twitter and 2012 for Mendeley. The absence of a significant difference in 2011 could be due to missing tweets, as altmetrics.com started accumulating altmetrics in that year. N/A values were mainly due to insufficient data.

Table 3.6 Statistical significance between NOA and OA articles across altmetrics and years

Altmetric type	**2010**	**2011**	**2012**	**2013**	**2014**
Blogs	0.04	0.43	0.44	0.14	0.60
CiteULike	0.00	0.00	0.00	0.00	0.96
F1000 reviews	0.00	0.00	0.00	0.00	0.60
Facebook	0.11	0.04	0.00	0.00	0.85
Google+	0.51	0.83	0.05	0.05	0.94
Mendeley	0.00	0.00	0.51	0.00	0.00
News outlets	0.46	0.04	0.00	0.00	0.00
Peer review sites	0.62	0.24	0.15	0.63	0.64
Reddit	0.10	0.75	0.73	0.51	0.53
Twitter	0.00	0.05	0.00	0.00	0.00
Weibo	N/A	0.53	0.49	0.64	0.01

We checked for significant differences between NOA and OA articles for journals and publication years on platforms that showed OAAA. Table 3.7 presents an example from Mendeley, which shows a significant difference for eight journals in 2014 but for only two in 2010. This could be because OA articles are available as preprints earlier than NOA articles, whereas in 2011 and 2012 only three and two journals, respectively, showed significant differences. In other platforms, we found similar results for journals showing a significant difference in 2014. However, we found less significant differences within years and journals overall.

Table 3.7 Statistical significance between NOA and OA articles for readership across journals and years

Journal name	2010	2011	2012	2013	2014
Accounts of Chemical Research	0.10	0.29	0.25	0.52	0.85
Advanced Materials	0.05	0.03	0.30	0.10	0.00
American Economic Review	N/A	N/A	N/A	N/A	0.02
American Journal of Respiratory and Critical Care Medicine	0.45	0.03	0.00	0.43	0.01
Astronomy and Astrophysics	0.97	0.99	0.43	0.85	0.59
Circulation	0.35	0.35	0.44	0.00	0.05
Clinical Infectious Diseases	N/A	N/A	N/A	0.00	0.00
European Heart Journal	0.35	0.01	0.01	0.00	0.14
Gastroenterology	0.00	0.68	0.23	0.00	0.70
Genes and Development	N/A	N/A	0.59	0.21	0.00
Hepatology	N/A	N/A	N/A	0.00	0.00
Journal of Clinical Oncology	0.99	0.68	0.72	N/A	0.09
Journal of Immunology	0.55	N/A	N/A	N/A	0.27
Journal of the American College of Cardiology	0.00	0.64	0.77	0.23	0.02
Neuron	N/A	N/A	N/A	0.01	0.00
Review of Financial Studies	0.23	0.85	0.40	0.07	0.81

Finally, we compared NOA and OA articles to determine whether there was a correlation between citation count and Mendeley readership, as shown in Figure 3.10. We selected articles published in 2012 to ensure that enough time had passed for them to accumulate citations and readership. We found a weak significant correlation between citation count and average readership for NOA articles ($\rho = 0.26$). However, we found a moderately significant correlation between citation count and average readership for OA articles (ρ = 0.56). No correlation was found between readership for NOA articles and readership for OA articles. Further, articles that received more than 80 citations were mostly OA with a significant difference, which shows a preference for sharing OA articles over NOA articles in academic social networks.

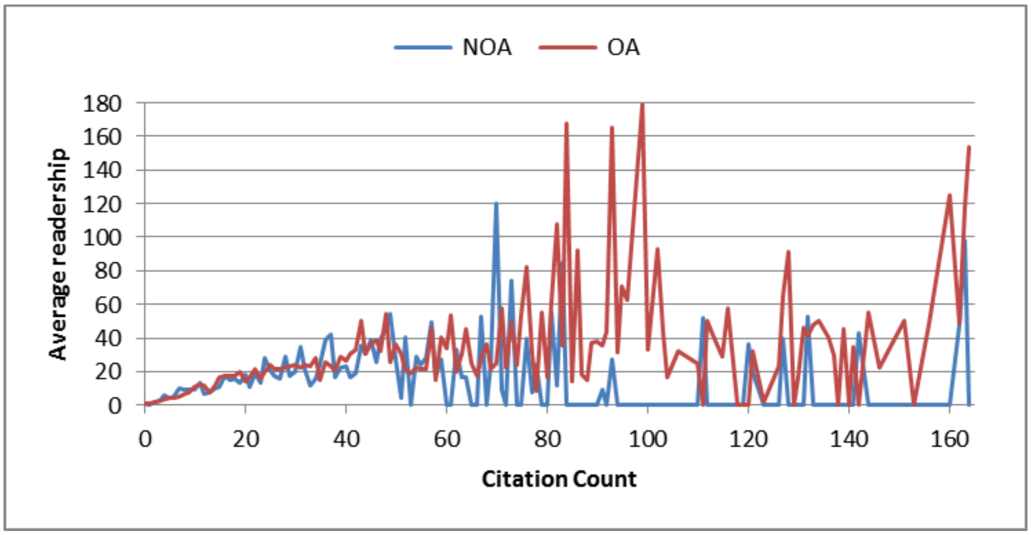


Figure 3.10 Average Mendeley readership per citation count for NOA and OA articles

4. LEVERAGING DATA FROM ACADEMIC SOCIAL NETWORKS TO IDENTIFY THE SOCIAL IMPACT AND PREDICT THE SCHOLARLY IMPACT[*]

4.1. Introduction

Rankings play a vital role in daily life. Students use rankings to decide which universities to apply to, patients use rankings to select hospitals, and travelers use rankings to plan vacations. Similarly, rankings of scholarly articles and venues are often used in academic and other research settings. The top scholarly venues have a great influence on research. Prestigious journals use rankings in their publicity, librarians refer to them when making decisions on subscriptions, researchers use them when determining where to submit their articles for publication, and research institutes use them in academic hiring, promotions, and funding decisions.

Journal rankings may not represent real research outcomes, as even low-ranked journals could publish good work. And, although concerns and objections have often been raised pertaining to such rankings, particularly in terms of their use in determining appointments, promotions, and research grants, they continue to be used. Arts and humanities scholars have raised additional concerns about whether the various rankings

[*] Part of this section is reprinted with permission from “Identifying the Real-time impact of the Digital Humanities using Social Media Measures” by Hamed Alhoori and Richard Furuta, 2013. Proceedings of the Digital Humanities Conference, 2013. Part of this section is reprinted with permission from “Can Social Reference Management Systems Predict a Ranking of Scholarly Venues?” by Hamed Alhoori and Richard Furuta, 2013. Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries, Copyright 2013 Springer-Verlag Berlin Heidelberg.

accommodate differences in culture, region, and language. According to Di Leo [203], "journal ranking is not very useful in academic philosophy and in the humanities in general" and one reason for this is the "high level of sub-disciplinary specialization." Additionally, Di Leo notes that there is "little accreditation and even less funding" in the humanities when compared with business and the sciences.

In a *Nature* article entitled "Rank injustice*," Lawrence* [204] notes that the "Impact factor causes damaging competition between journals since some of the accepted papers are chosen for their beneficial effects on the impact factor, rather than for their scientific quality." Another concern is the effect on new fields of research. According to McMahon [205], "Film studies and media studies were decimated in the metric because their journals are not as old as the literary journals. None of the film journals received a high rating, which is extraordinary."

Although the Australian government dropped rankings after complaints that they were being used "inappropriately," it still offers a profile of journal publications that provides an "indication of how often a journal was chosen as the forum of publication by academics in a given field" [206]. Despite concerns over rankings, educators and researchers agree that some kind of quality management system is necessary. By publishing their results, researchers are not just talking to themselves. Research outcomes are for public use, and others should be able to study and measure them.

The impact of scholarly articles and venues is typically measured using citation analysis. However, it can be months or even years before the importance of an article can be determined based on citations. Moreover, research articles, especially those published in conference proceedings, are limited in terms of length, such that authors may not cite all the related references.

An alternative approach to citation analysis is that of using data from online scholarly social networks [158]. Scholarly communities have used social reference management (SRM) systems to find, store, share, and discover scholarly articles and references [19]. By storing articles and references online, researchers can archive their research interests without encountering any limits. Therefore, the statistics for these online repositories are strong indicators of researchers' interests and may reflect research interests more accurately than statistics about downloads or views.

In this section, we attempt to answer the central questions:

- How can we measure research efforts and their impact? And, how can we get an early indication of research work that is capturing the research community's attention?
- Are measures appropriate for one research area also applicable to publications in a different area?
- Can we predict a scholarly venue ranking using social-based metrics?
- What is the effect of open-access venues on rankings?

We seek insights into ways to answer these questions by using data from a social media site to measure the real-time impact of articles in the digital humanities and to predict rankings of scholarly venues. This section is structured as follows: We discuss related research in Subsection 4.2. Following in Subsection 4.3, we describe the experiments, data collection, and methodology. In Subsection 4.4, we present and discuss our results.

4.2. Related Work

Although the controversial "impact factor" is a well-known method for ranking scholarly venues, it suffers from citation delay [177], differs according to discipline [207], and may not be available for emergent venues. The Science Journal Ranking (SJR) indicator [208], which takes into account the quantity and quality of the citations, has been proposed as an alternative to the impact factor. A number of journal-ranking approaches use the PageRank algorithm, including the SJR indicator and the Eigenfactor [209]. The h-index, expert survey [210], and publication power approach [211] have also been used to rank venues.

The research on ranking academic venues is relatively extensive and wide-ranging. For example, Zhuang et al. [212] used program committee characteristics to discover and rank conferences. Yan et al. [213] defined two approaches to ranking academic venues: a seed-based approach drawing on author meta-data and a browsing-based approach drawing on both citation and author meta-data. Martins et al. [214] used a large number of features with machine learning techniques to assess the quality of scientific conferences. Rahm et al. [215] found that conferences could have a higher impact factor than journals. Google Scholar joined the effort to rank venues by launching Scholar Metrics, which ranks top scholarly venues in several disciplines and languages, on the basis of the five-year h-index.

Li et al. [216] compared the Web of Science citation counts and the CiteULike/Mendeley readership counts on a limited sample of articles published in *Nature* and *Science* and found significant correlations between the two rankings. Kraker et al. [217] found a significant relationship between Mendeley references and the SCImago's impact index, which is SCImago's version of the impact factor. They also found differences among disciplines.

4.3. Data and Experiments

4.3.1. Research Community Article Rating (RCAR)

Tenopir and King [218] estimated that scientific articles published in the U.S. are read about 900 times each. Who are the researchers reading any given article? Does knowing who these researchers are influence the article's impact? Rudner et al. [219] used a readership survey to determine the researchers' needs and interests. Eason et al. [220] analyzed the behavior of journal readers using logs.

There is a difference between how many times an article has been cited and how many times it has been viewed or downloaded. A citation refers to an instance in which an author has probably read the article, although this is not necessarily the case. In respect to article views, there are several viewing scenarios such as intended clicks, unintended clicks, or even a web crawler. Therefore, the number of views has hidden influential factors. To eliminate the hidden-factors effect, we selected articles that researchers had added to an academic social media site. In this study, we ranked readers based on educational level. For example, a professor had a higher rank than a PhD student, who in turn had a higher rank than an undergraduate student.

Zotero's readership statistics were not available to the public, and in CiteULike, the most cited articles in *Literary and Linguistic Computing* (*LLC*) were shared by few users. Therefore, we were unable to use either system's data. Instead, we obtained our data from Mendeley, using its API [221]. We measured the research community article rating (RCAR) using Equation 4.1:

$$RCAR = \frac{\sum R + \sum(P * K) + \sum D + \sum C + \sum A + \sum G}{\log(y_c - y + 2)} \tag{4.1}$$

RCAR uses the following measures:

- R = researchers who added an article to their online profiles in an academic social network
- $\sum(P * K)$ = percentage (P) of researchers who added an article, multiplied by their rankings (K)
- $\sum D$ = number of academic disciplines represented by R
- $\sum C$ = number of countries represented by R
- $\sum A$ = number of authors credited on an article
- $\sum G$ = number of online groups that shared an article
- y_c = current year
- y = year the article was published

4.3.2. Scholarly Venues

We crawled CiteULike and downloaded 554,023 files, in which each file includes a reference to an article and a list of the users who added the article to their profiles. We used only files that included details about either a conference or a journal, for a final sample of 407,038 files. We then extracted the details of the venues and collected a total of 1,317,336 postings of researcher–article pairs and a total of 614,361 researcher–venue pairs. We defined three social-based metrics and used them to rank venues:

1. **Readership:** The number of researchers who added references from a venue to a social reference management system.
2. **Article Count:** The number of unique articles from a single venue added to an SRM system.
3. **Active Researchers Rating (ARR):** We defined active researchers as those who added twenty or more venues to their online repositories. We used a weighted sum to increase the importance of newly added references. Equation 4.2 was used to compute the ARR for venue v:

$$ARR(v) = \sum_{i=1}^{n} \sum_{w=m}^{1} w \log(v_w + 1) \tag{4.2}$$

The outer summation of the ARR totals the individual ratings for n researchers. In the inner summation, v_w denotes the number of references from a specific venue that a researcher added to his/her profile in a given year, out of all the m years during which the researcher followed venue v. Weight w increased the importance of newly added references. The ARR favors researchers who followed venues for several years over researchers who added numerous references from venues for a few years. The log minimized the effect of adding large numbers of references.

We compared the Google Scholar h5-index with our social-based rankings. Currently, Google Scholar h5-index includes research articles published between 2007 and 2011 and indexed in Google Scholar as of November 2012. To compare our social-based rankings with Google Scholar h5-index, we selected articles published and added to CiteUlike between 2007 and 2011. Our question was whether a correlation exists between social metrics from CiteULike and citation metrics from Google Scholar h5-index for the indicated time span. We repeated this strategy with the other citation-based rankings. For example, the Eigenfactor score, which relies on Web of Knowledge citations, was released in 2011 and includes articles published between 2006 and 2010. Therefore, in this instance, we used a dataset of articles that had been published and added to CiteULike between 2006 and 2010.

We used Spearman's rank correlation coefficient, ρ(rho), to compare our social-based rankings with a number of citation-based rankings, such as the Google Scholar h5-index, the SCImago h-index, the Thomson Reuters Impact Factor, the Eigenfactor score, and the total number of citations. We began with citation-based rankings and mapped the corresponding values from the social-based rankings.

4.4. Results and Discussion

4.4.1. Scholarly Articles

4.4.1.1. Citations, readership, and RCAR

We looked at seven digital humanities journals included in Mendeley and mentioned on Wikipedia [222]. Of these seven journals, only two had an h5-index on Google Scholar: *Digital Creativity* (h5-index = 7) and *LLC* (h5-index = 13). We calculated the RCAR and compared the top-cited *LLC* articles based on the number of Google Scholar citations and the number of Mendeley readership, as shown by sample articles in Table 4.1. The number of citations was significantly higher than the number of Mendeley readership for *LLC* (p-value < 0.05).

Table 4.1 Google citations, Mendeley readership, and RCAR for LLC articles

Article title	Citations	Readership	RCAR	Year
Quantitative Authorship Attribution: An Evaluation of Techniques	73	42	84.85	2007
If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data	37	16	46.03	2008
An evaluation of text classification methods for literary study	32	16	40.97	2008
Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts	35	17	41.11	2007
All the Way Through: Testing for Authorship in Different Frequency Strata	25	11	29.24	2007
Function Words in Authorship Attribution Studies	28	16	39.37	2007
Use of the Chi-Squared Test to Examine Vocabulary Differences in English Language Corpora Representing Seven Different Countries	24	9	27.05	2007
Interpreting Burrows's Delta: Geometric and Probabilistic Foundations	23	13	38.10	2008
Supporting Annotation as a Scholarly Tool—Experiences From the Online Chopin Variorum Edition	19	20	48.10	2007
Modelling Space and Time in Narratives about Restaurants	20	8	22.35	2007
Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification	21	17	44.91	2008
The Identification of Spelling Variants in English and German Historical Texts: Manual or Automatic?	16	9	31.52	2008
The effect of author set size and data size in authorship attribution	20	18	81.24	2011

We investigated ways in which the discipline of digital humanities differs from other disciplines. We compared *LLC* with a journal from a different area of research, *Library Trends,* which had a similar h5-index. *Library Trends* received more citations and readership than *LLC*. Three of the top articles in *Library Trends* also had more Mendeley readership than citations, whereas this was the case for only one *LLC* article. However, there was no significant difference between *Library Trends* citations and readership. Next, we tested the *Journal of the American Society for Information Science and Technology* (JASIST) and the *Journal of Librarianship and Information Science* (JOLIS). We found that JASIST and JOLIS readership of articles published in 2012 were significantly higher than the citations. This indicates that computer, information, and library scientists are more active in academic social networks than digital humanities researchers. By active, we mean that these researchers share and add more newly published articles to their online repositories.

4.4.1.2. Citations and altmetrics

In order to better understand various socially based measures, we used altmetrics and citations to compare *LLC* articles. We used an implementation of a*ltmetrics* whereby, altmetrics score "each article [receives] a score that measures the quantity and quality of attention it has received from Twitter, Facebook, science blogs, mainstream news outlets and more sources" [223]. We found that most of the articles that received social media attention were published during the last two years. However, a number of articles that were published four or more years ago constituted exceptions to this finding.

These older articles received at least four citations, as shown in Table 4.2. We also found similar correlations for articles in *Digital Creativity*.

Finally, we compared the LLC articles on the basis of readership and altmetrics score. We found no significant difference between LLC citations of articles published in 2012 and readership. However, we found a significant difference between altmetrics score and citations ($p < 0.05$) for articles published in 2012. This shows that researchers interested in digital humanities are more active on general social media sites (e.g., Twitter and Facebook) than on academic social media sites (e.g., Mendeley).

Table 4.2 Altmetrics score and citations for LLC articles

Article	Altmetric	Citations	Year
Transcription maximized; expense minimized? crowdsourcing and editing The Collected Works of Jeremy Bentham	17.55	2	2012
Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists	12.45	4	2011
A rationale of digital documentary editions	6.45	4	2011
Computational analysis of the body in European fairy tales	6.3	1	2012
Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification	5.35	22	2008
Experiments in 17th century English: manual versus automatic conceptual history	4.35	0	2012
Improving record matching in imprecise and uncertain datasets	3.75	0	2012
Managing and Growing a Cultural Heritage Web Presence. A strategic guide. Mike Ellis.	3.25	0	2012
Natural language processing and early-modern dirty data: applying IBM Languageware to the 1641 depositions	2.75	0	2012
Scalability Issues in Authorship Attribution.Kim Luyckx.	2.75	4	2011
Detecting authorship deception: a supervised machine learning approach using author writeprints	2.25	1	2012
Co-occurrence-based indicators for authorship analysis	2	0	2012
A thing not beginning and not ending': using digital tools to distant-read Gertrude Stein's The Making of Americans	2	13	2008
It's a team if you use "reply all" ': An exploration of research teams in digital humanities environments	2	15	2009
Who wrote Shamela? Verifying the Authorship of a Parodic Text	2	4	2005
The Density of Latinate Words in the Speeches of Jane Austen's Characters	1.85	9	2001
The inadequacy of embedded markup for cultural heritage texts	1.85	9	2010
Visual Interface Design for Digital Cultural Heritage. A Guide to Rich-Prospect Browsing. Stan Ruecker, Milena Radzikowska, and Stéfan Sinclair (eds).	1.75	0	2012
The Tesserae Project: intertextual analysis of Latin poetry	1.75	0	2012
Ce qui compte. Méthodes statistiques. Ecrits choisis, tome II. Etienne Brunet (edited by Céline Poudat).	1.75	0	2012
The Potosi principle: Religious prosociality fosters self-organization of larger communities under extreme natural and economic conditions	1.75	0	2012
Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviour, Products and Institutions. Willard McCarty (ed).	1.6	0	2012
How To Do Things With Videogames. Ian Bogost.	1.6	0	2012
Digital Research in the Study of Classical Antiquity.Gabriel Bodard and Simon Mahony.	1.6	0	2012
In Memoriam Charles Douglas Bush (1948-2011)	1.6	0	2011
A trip around the world: Accommodating geographical, linguistic and cultural diversity in academic research teams	1.5	0	2012
Poetics of crisis or crisis of poetics in digital reading/writing? The case of Spanish digital literature	1.25	0	2012
Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project	1	10	2008
Narrative rules? Story logic and the structures of games	1	0	2012

4.4.2. Scholarly Venues

First, we compared the general citation-based rankings of the top 100 venues with our social-based rankings and found strong positive relationships ($p < 0.01$), as shown in Table 4.3. There was no significant correlation between the social-based metrics and the impact factor or the impact index.

Table 4.3 Correlations between citation-based metrics and social metrics for the top 100 venues

Citation-based metric	Readership	ARR	Article count
SCImago h-index	0.581	0.566	0.534
Google h5-index	0.336	0.354	0.349
Eigenfactor score	0.688	0.669	0.665
Total citations	0.675	0.625	0.632

We then compared the top 20 venues among various research areas using Google's h5-index and social-based metrics. We found significant relationships in some areas, as shown in Table 4.4. In Tables 4.4 and 4.5, we used * to represent ($p < 0.05$) and ** to represent ($p < 0.01$). We also compared Google Scholar h5-index with the social metrics for some sub-disciplines in engineering and computer science, as shown in Table 4.5.

Table 4.4 Correlations between the Google Scholar 5h-index and social metrics for various research areas

Research area	Readership	ARR	Article count
Health & medical sciences	0.647 **	0.672**	0.642**
Humanities, literature & arts	0.368	0.471	0.200
Life sciences & earth sciences	0.788 **	0.768 **	0.735 **

Table 4.5 Correlations between the Google Scholar 5h-index and social metrics for some engineering and computer science sub-disciplines

Sub-discipline	Readership	ARR	Article count
Automation & control theory	0.567 *	0.382	0.466
Bioinformatics & computational biology	0.814 **	0.700 **	0.706 **
Educational technology	0.575 *	0.512 *	0.374
Library & information science	0.761 **	0.769 **	0.754 **
Robotics	0.532 *	0.482	0.460 *

No significant relationships were found between the Google Scholar h5-index and the social-based rankings in some areas, such as arts and humanities. This was also the case for some sub-disciplines, such as artificial intelligence. However, we found a significant relationship between the SCImago h-index and the readership ranking in arts and humanities ($p < 0.05$) and in artificial intelligence ($p < 0.01$). Surprisingly, and in most cases when compared with the citation-based rankings, the readership rankings had higher correlations than did the ARR. The article count usually had weaker correlations than readership and ARR.

As shown in Table 44.3, it is clear that social metrics are an effective way to measure the popularity of venues because such metrics have a strong positive correlation with the total number of venue citations. Social metrics can also measure the quality of venues, as they are strongly positively correlated with quality ranking methods, such as Eigenfactor scores. Tables 4.4 and 44.5 show differences in correlations among various research areas—differences could be due to varying levels of online scholarly activity. Moreover, such differences may also relate to unequal distributions of research communities across SRM systems or to the existence of research communities that are not active in such online systems. We experimented with two social-based metrics that resemble the impact factor, but we did not find any strong correlation. For the first metric, we divided the readership of a venue by article count, and for the second metric, we divided the ARR by article count.

Finally, we investigated whether the venue-ranking approach (citation-based or social-based) was related to type of access to venues (subscription or open access). We also compared the top 20 venues in the Google h5-index with the top 20 venues in readership and ARR rankings. We included hybrid and delayed access venues in the open-access venue category. There were more open-access venues in the readership and ARR rankings than in the citation-based rankings. We did not find a significant relationship for the readership ranking. However, using the ARR, we found 13 open-access venues but only 6 in the Google h5-index. And, a Chi-squared test determined that there was a significant positive relationship ($X^2 = 4.9123$, $p < 0.05$) between the venue-ranking approach and type of access to venues.

5. RATING AND RECOMMENDING SCHOLARLY VENUES BASED ON TEMPORAL ANALYSIS OF PERSONAL ALTMETRICS

5.1. Introduction

Generally, researchers become aware of scholarly venues related to their research interests by word of mouth from lab members, departmental colleagues, members of other scholarly communities, through conducting online searches and reviewing research articles they come across, from rankings of venues, and publishers' reputations [37][224]. Earlier approaches worked satisfactorily, as there were relatively few venues related to any given field. However, in today's more multifaceted scholarly environment, researchers can only become acquainted with newly available and specialized venues by spending considerable time browsing and evaluating.

It is also essential for funding agencies to become cognizant of new lines of research across fields in order to determine plans for future funding. Further, new cross-over research areas lead to more 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, whereas, too often, the venue's impact factor and/or users' requests are emphasized. For example, Springer provides its users with a form for recommending journals to librarians [225], but this feedback represents individual interests rather than providing a picture of the entire constituency's needs.

Rankings of scholarly venues have been created and used to help researchers become aware of specific scholarly communities. However, knowing that prestigious journals, such as *Science* and *Nature*, are top venues for multidisciplinary fields is not useful to researchers seeking more specialized venues and communities. Moreover, traditional citation analysis does not provide real-time results, especially for new scholarly venues, that do not have an impact factor.

A number of online services provide collections or notifications of venues. For example, the HCI Bibliography [226] is a specialized bibliographic database on Human-Computer Interaction. AllConferences[2] and Lanyrd[3] are global conference and events directories. ConferenceAlerts,[4] EventSeer,[5] and WikiCFP[6] provide notifications of upcoming academic events based on keywords. ConfSearch [227] enable researchers to search for computer science conferences using keywords, related conferences, and authors.

In this era of big data, retrieving relevant results by searching and browsing online is no longer the only approach nor it is necessarily the most efficient way. Studies have been conducted in an effort to offer techniques capable of accelerating scholarly discovery, such as summarization, visualization [228], and collaborative information synthesis [229]. Further, recommender systems have been introduced to filter the overwhelming amount of data by using various techniques to alleviate information

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

overload [230][231]. Currently, recommender systems provide millions of online users with continually updated suggestions for news, books, restaurants, vacation packages, and movies.

With the proliferation of publications, researchers are utilizing academic social networks and reference management systems in order to find, store, and manage references [19]. Social or online reference management systems enable users to bookmark references to research content, consisting mainly of research articles. These tools enable users to tag, review, and rate research content within their profiles. Such scholarly tools play an essential role in the organization of personal article collections and generation of bibliographies. Research groups have been formed, and scholarly communities are sharing their digital collections of references. Such online personal collections or repositories reflect researchers' current reading and indicate changes in their interests over time.

In Section 2, we found that several of the researchers who participated in the studies express a desire to be aware of new and well-established scholarly venues and events related to their shifting research interests. In this section, we build a personal measure for evaluating venues based on user-centric altmetrics and 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,[7] a well-known social reference management system.

This section is structured as follows: In Subsection 5.2, we discuss related work. In Subsection 5.3, we describe an approach for measuring an implicit rating for scholarly venues by monitoring researchers' behavior. In Subsection 5.4, we explain the data collection and the experiments. In Subsection 5.5, we present and discuss the results.

5.2. Related Work

5.2.1. Recommending Venues

A few studies have been conducted on recommending scholarly venues. Klamma et al. [232] recommended academic events based on researchers' event participation history. Luong et al. [233] used co-authors' publication history to recommend venues. Boukhris and Ayachi [234] proposed a hybrid recommender for upcoming conferences in computer science based on venues from co-authors, co-citers, and co-affiliated researchers. Pham et al. [235] 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. Other venue recommendation approaches based their ratings on the topic and writing style of a paper [236], the title and abstract of a paper [237], and personal bibliographies and citations [238][239].

In addition, research has been carried out on recommending events in general. For example, Minkov et al. [240] proposed an approach to recommending future events,

[7] http://www.citeulike.org/

whereas Khrouf and Troncy [241] used hybrid event recommendations with linked data. Quercia et al. [242] used mobile phone location data to recommend social events.

Most research to date used citation analysis and the publication or participation history of researchers to recommend venues, which would not be useful for new researchers or graduate students without an established record of scholarly activity. Furthermore, using only the venues in which a researcher has previously published work, would undermine the recommendation process, as a researcher might be interested in new research areas in which s/he has not yet published. The present research study explores a way to draw on a researcher's current personal article collections and readings to recommend tailored venues.

5.2.2. Recommender Systems

Recommender systems augment the decision-making process without having adequate experience of the options [243]. One well-known recommending technique is collaborative filtering (CF) [244][245][246][247][248], which recommends items based on preferences from other similar users (user-based CF) or from similar ratings received by items (item-based CF [249]). CF has been used in several domains, including recommending movies [250], music [251], and books [252]. Another commonly used recommendation technique is content-based filtering [253], which recommends items similar to those a user has selected based on item descriptions or other user data, and is most widely used in textual domains [254]. CF is affected by the cold-start problem [255], in which the system cannot produce good recommendations for new users or unrated items. This problem can be remedied to some extent by using a hybrid approach that combines CF and content-based filtering [256] or by using pseudo-users who provide ratings according to the attributes of items or users [257]. However, CF has some important benefits, in that it provides recommendations for items that are complex to analyze and it occasionally provides serendipitous recommendations [258]. Other recommenders have used a matrix factorization approach based on the stochastic gradient descent (SGD) [259], singular value decomposition (SVD) [260], or SVD++ [261], which addresses the issues of sparsity and scalability.

Recommender systems have also been used in scholarly environments to recommend research papers, collaborators, reviewers, citations, and tags. Further, the processes whereby scholarly articles are recommended have been widely studied in recent years and applied to academic social networks [262][263]. Torres et al. [264] recommend research articles based on a hybrid approach that used citations, paper titles, and abstracts. Bogers et al. [20] experimented with three different CF algorithms to recommend papers using CiteULike and found that user-based filtering performed best. Sugiyama et al. [265] modeled a researcher's publications and the publications cited therein as a basis for recommending research papers. Agarwal et al. [266] used searches performed by researchers with similar interests to support existing search engines with recommendations, whereas Ohta et al. [267] proposed a scholarly browsing system augmented by recommending related papers. Beel [268] used mind maps to find relatedness and to recommend documents, and Guan et al. [269] proposed a graph-based representation of a learning algorithm for recommending documents using tags. Gori et al. [270] presented a research paper recommender system using the ACM dataset and a random walk algorithm. Pohl et al. [271] used digital access records (e.g., http-server logs) to recommend papers. Scienstein [272] is a hybrid recommender system for research papers that analyzes keywords, citations, authors, sources, and ratings, and Nascimento et al. [273] used terms present in papers to generate paper recommendations. OSUSUME [274] introduced a Japanese research paper recommender system.

Google Scholar released Scholar Updates [275], a research article recommender system that determines article relevance using a statistical model based on the researcher's published work, the citation graph, and the co-authors. Scholar Updates requires the creation of a Google Scholar public profile, as recommendations are restricted to authors and are based on their publications, but not their current reading. Figure 55.1 shows research article recommendations from Mendeley. Other scholarly article recommenders include CiteULike, Faculty of 1000 Prime,[8] ReadCube,[9] Sciencescape,[10] Sparrho,[11] PubChase,[12] and Scizzle.[13]

Research has been conducted to determine how best to recommend collaborators [276][277], experts [278], and reviewers [279][280]. Other uses of recommendation systems include citation recommenders [281][282][283][284][285][286] and tag recommenders [287][288][289][290].

---

[8] http://f1000.com/prime
[9] https://www.readcube.com/
[10] https://sciencescape.org/
[11] http://www.sparrho.com/
[12] https://www.pubchase.com/
[13] http://www.myscizzle.com/

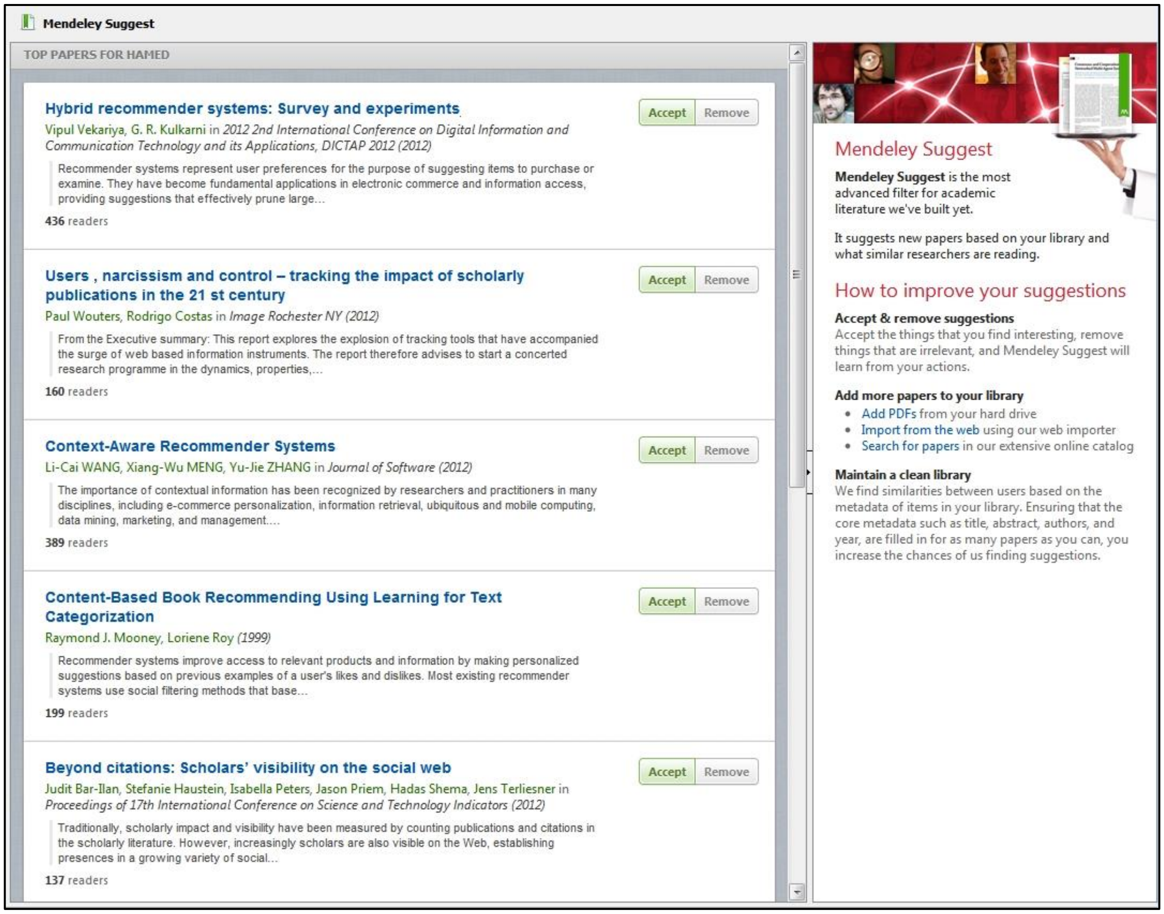


Figure 5.1 A screenshot of Mendeley article suggestions

5.3. Personal Venue Rating (PVR)

Research articles are associated with several metadata fields that can be used to produce recommendations. However, no direct metadata or ratings exist for venues. Nevertheless, references in a researcher's library can provide indirect information pertaining to a researcher's interests. We used references and the years in which they were added to a researcher's library as factors in the measurement of personal venue rating. PVR takes into consideration how a researcher's interest in a given venue has changed over time. In Equation 5.1, we define PVR as a weighted sum for researcher u and venue v, and we refer to it as $r_{u,v}$:

$$r_{u,v} = \sum_{i=y}^{1} w \log(v_{u,i} + 1) \tag{5.1}$$

$v_{u,i}$ denotes the number of references in a researcher's u library from a specific venue v, which the researcher added during a certain year of the total number of y years 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 shilling attempts [291]. 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 personalized than the published year.

5.4. Data and Experiments

5.4.1. Metrics

We conducted an offline experiment using our CiteULike dataset, collected as described in Section 3. We used user-based CF, item-based CF, SGD, and SVD++ algorithms from the Apache Mahout [292] to recommend venues to researchers. We compared researchers with similar interests in terms of their PVRs. To identify similarities among the researchers, we used the cosine similarity, the Pearson correlation similarity, and the Euclidean distance similarity [258].

The cosine similarity ($sim_{x,u}$) between a researcher x and another researcher u was computed as Equation 5.2, where $\vec{x}$ and $\vec{u}$ are two vectors representing the ratings of the two researchers, and the cosine similarity is the cosine angle between them:

$$sim_{x,u} = \cos\ (\theta) = \frac{\vec{x}\ \cdot\ \vec{u}}{||\vec{x}||\ \times ||\vec{u}||} = \frac{\sum_{v=1}^{n}\left(r_{x,v}\right)\left(r_{u,v}\right)}{\sqrt{\sum_{v=1}^{n}\left(r_{x,v}\right)^2}\ \ \sqrt{\sum_{v=1}^{n}\left(r_{u,v}\right)^2}} \tag{5.2}$$

$||\vec{u}||$ is the vector's Euclidian length, and n is the number of venues rated by both researchers. The Pearson correlation similarity ($sim_{x,u}$) is measured by Equation 5.3:

$$sim_{x,u} = \frac{\sum_{v=1}^{n}\left(r_{x,v} - \bar{r_x}\right)\left(r_{u,v} - \bar{r_u}\right)}{\sqrt{\sum_{v=1}^{n}\left(r_{x,v} - \bar{r_x}\right)^2}\ \ \sqrt{\sum_{v=1}^{n}\left(r_{u,v} - \bar{r_u}\right)^2}} \tag{5.3}$$

$\bar{r_u}$ is the average PVR for researcher u. Equation 5.4 shows the Euclidean distance:

$$Euclidean\ distance(x,u) = \sqrt{\frac{\sum_{v\,\in\,V_{x,u}}(r_{x,v} -\ r_{u,v}\ )^2}{|V_{x,u}|}} \tag{5.4}$$

$V_{x,u}$ is the set of venues rated both by x and u.

In the Euclidian distance similarity, a larger distance indicates fewer similar researchers; therefore, we used $(1/(1 + distance)$ to identify similar researchers. To decrease the importance of a few co-rated venues that would otherwise have created high correlations between active researchers, we applied a significance weighting [258].

Users tend to assign a certain range of ratings, such that some users may generally assign high ratings whereas others generally assign low ratings. Therefore, we normalized the ratings using a user mean-centering prediction [258]. Prediction $p_{x,v}$ for an active user x and for venue v is measured by Equation 5.5:

$$p_{x,v} = \overline{r_x} + \frac{\sum_{u \in U_v(x)} \left(r_{u,v} - \overline{r_u} \right) sim_{x,u}}{\sum_{u \in U_v(x)} |sim_{x,u}|} \tag{5.5}$$

$\overline{r_x}$ 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. $\overline{r_u}$ is the average rating for user u for the items rated by both x and u (i.e., all the co-rated items).

We also calculated the item mean-centering prediction, as shown in Equation 5.6:

$$p_{x,v} = \overline{r_v} + \frac{\sum_{w \in W_x(v)} \left(r_{x,w} - \overline{r_w} \right) sim_{v,w}}{\sum_{w \in W_x(v)} |sim_{v,w}|} \tag{5.6}$$

$\overline{r_v}$ 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). $\overline{r_w}$ is the average rating for venue w derived from the ratings of all the users who rated venues w and v.

5.4.2. 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 [293]. To rank the Boolean recommendations, venues affiliated with a large number of similar users were weighted more heavily [294].

To measure the recommendations' performance, we used precision, recall, and normalized discount cumulative gain (NDCG) [295][296]. Precision is derived by dividing the number of relevant venues recommended according to the researcher's venues by the number of recommended venues, as shown in Equation 5.7. Recall is derived by dividing the number of relevant venues recommended by the number of relevant venues, as shown in Equation 5.8. 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).

$$Precision = \frac{|relevant\ venues \cap top\ venues|}{|top\ venues|} \tag{5.7}$$

$$Recall = \frac{|relevant\ venues \cap top\ venues|}{|relevant\ venues|} \tag{5.8}$$

Discounted cumulative gain (DCG) measures the extent to which a venue ranking is relevant to a user's ideal ranking, as shown in Equation 5.9:

$$DCG_p = \sum_{v=1}^{p} \frac{2^{rel_v} - 1}{log_2(1+v)} \tag{5.9}$$

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 Equation 5.10:

$$NDCG_p = \frac{DCG_p}{IDCG_p} \tag{5.10}$$

As recommendation lists vary in length, we used NDCG. $IDCG_p$is the maximum possible ideal DCG at position *p*.

We also incorporated user coverage [297][298][299], 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 [300], the absolute deviation of a researcher's predicted PVR and observed PVR, is calculated as shown in Equation 5.11:

$$MAE = \frac{\sum_{v=1}^{n} \left| p_{u,v} - r_{u,v} \right|}{n} \tag{5.11}$$

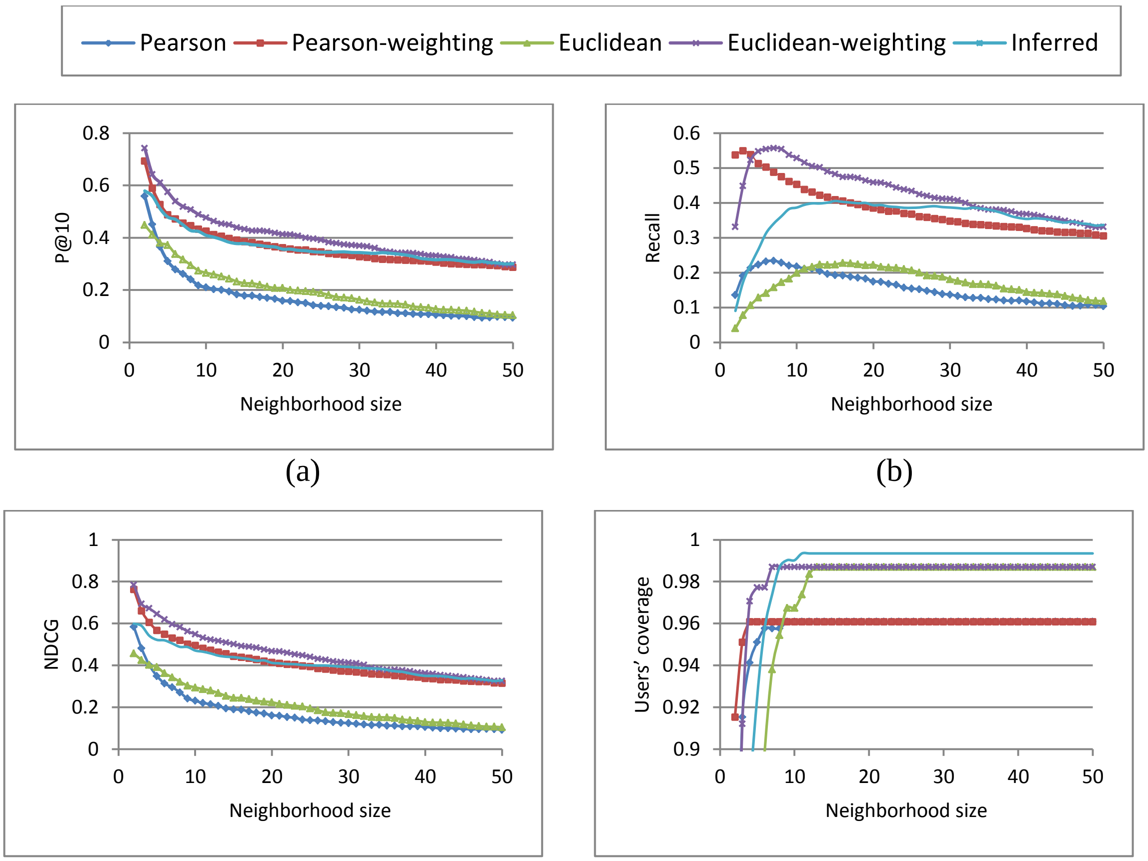
RMSE is measured using the square root of the average squared difference between a researcher's predicted PVR and observed PVR as shown in Equation 5.12:

$$RMSE = \sqrt{\frac{\sum_{v=1}^{n} (p_{u,v} - r_{u,v})^2}{n}} \tag{5.12}$$

$p_{u,v}$ is the predicted rating for venue v, and $r_{u,v}$ is the actual rating. 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.

5.5. Results and Discussion

We began by comparing user similarities with and without significance weighting. Inferred ratings, i.e., is the average researcher's ratings, were used for venues that researchers did not rate. Figure 5.2 shows that using significance weighting improved the accuracy, recall, and NDCG. Using inferred ratings showed some improvement in the results as the neighborhood size increased.



(a) (b) (c) (d)

Figure 5.2 A comparison of user-based CF algorithm with different similarities and neighborhood sizes

We then compared similarities that used PVR ratings and the user-based CF algorithm with the Boolean recommendation, i.e., the baseline, as shown in Figure 5.3. Figure 55.3 (a–c) demonstrates that in general, the PVR implicit ratings achieved higher precision, recall, and NCDG at lower neighborhood sizes. Figure 5.3 (d) also shows the users' coverage and that the PVR provided recommendations for up to 98% of users.

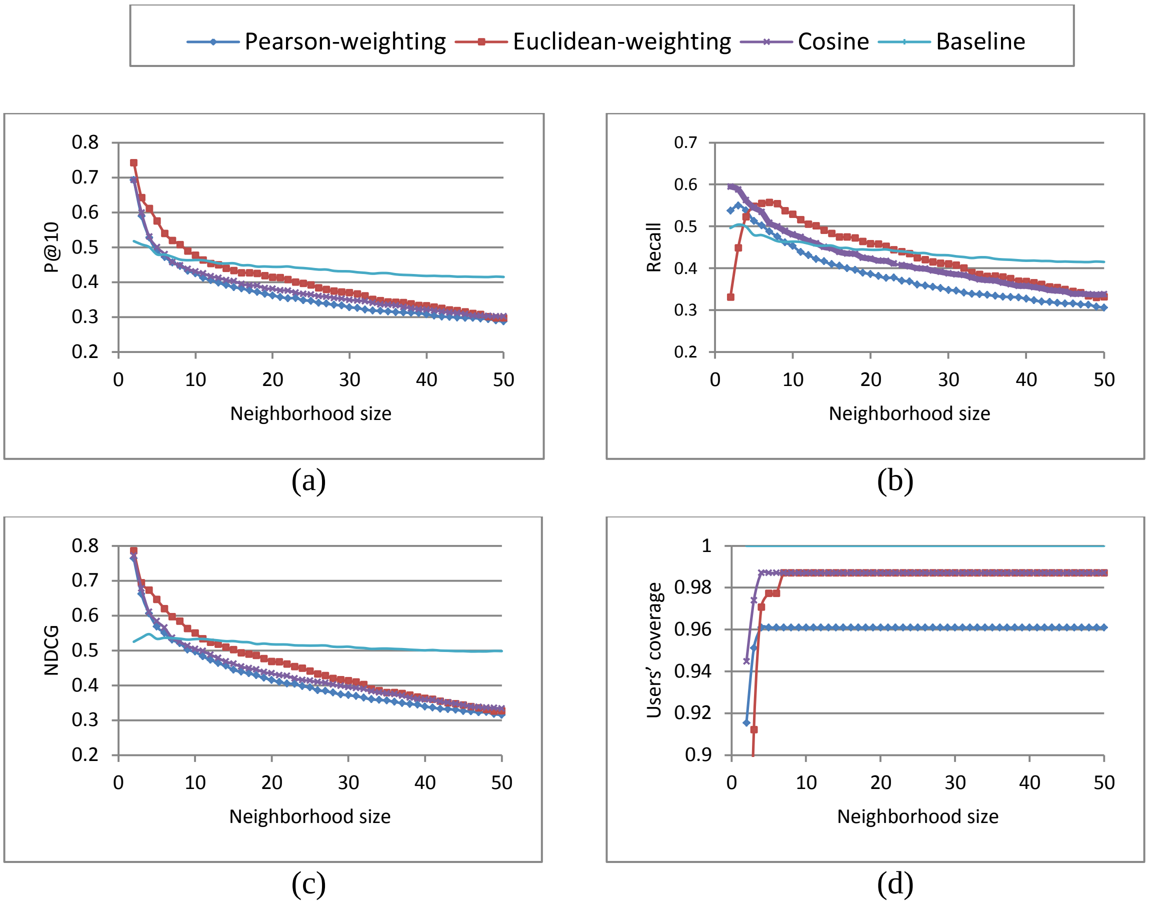


Figure 5.3 A comparison of user-based CF algorithm with similarities that use PVR ratings and the baseline at different neighborhood sizes

Figure 5.4 illustrates the use of thresholds for users instead of fixed neighborhood sizes. Pearson-weighting achieved the highest P@10 and the highest NDCG, whereas Boolean recommendations achieved the highest recall and the highest coverage.

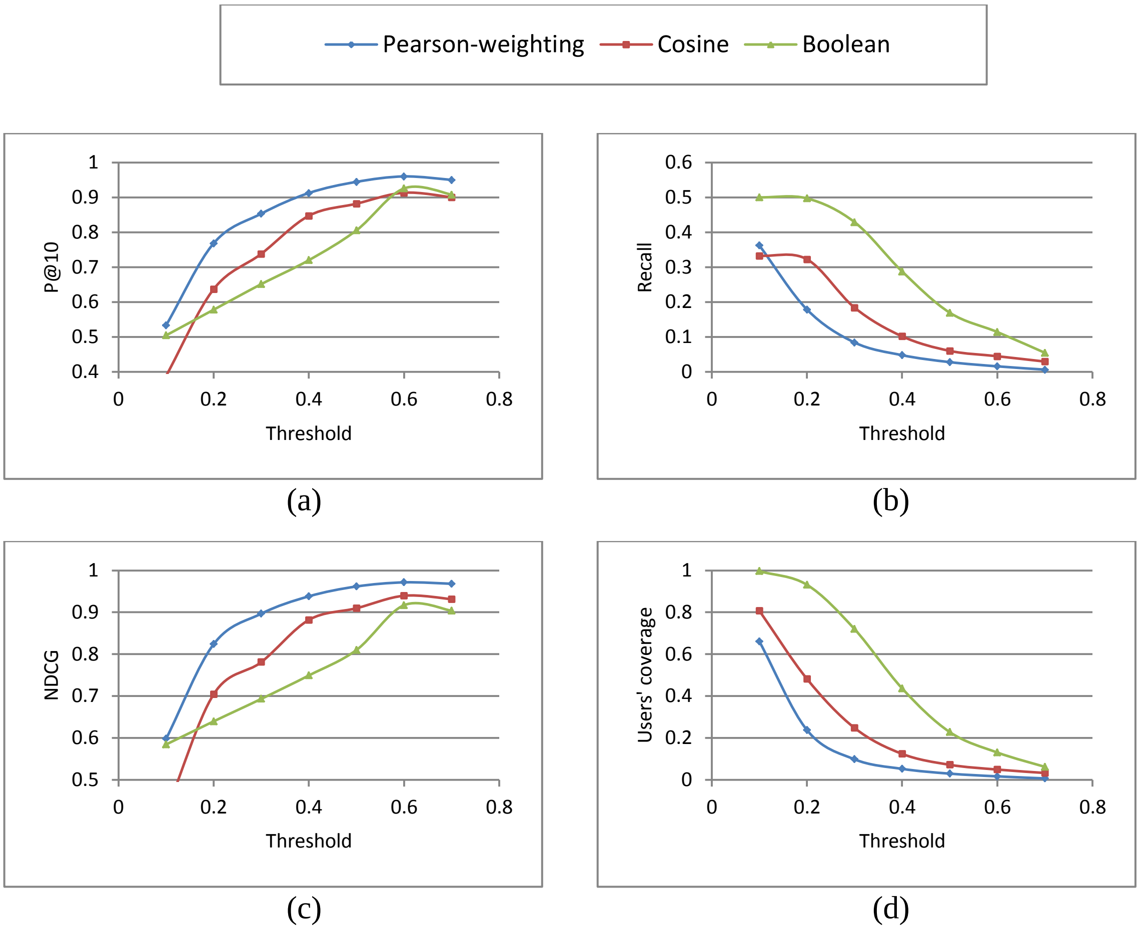
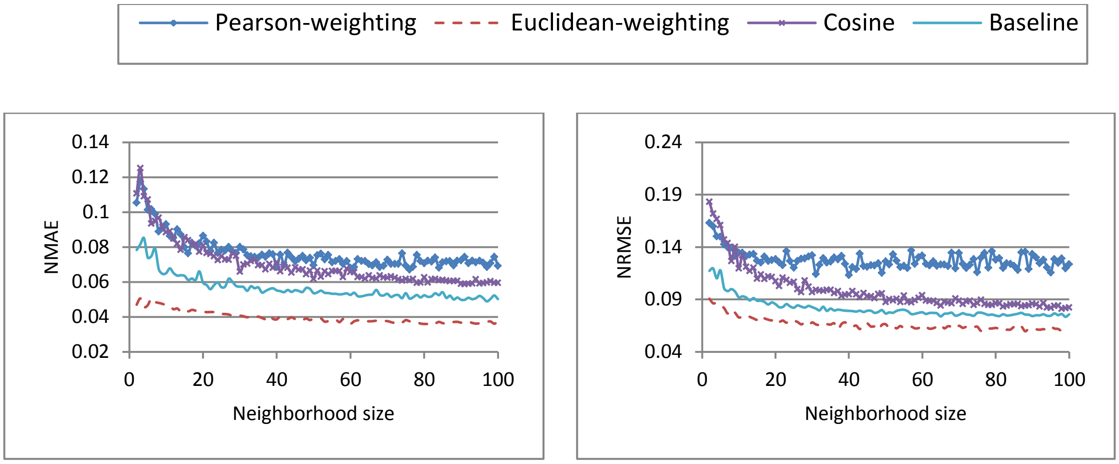


Figure 5.4 A comparison of user-based CF performance using different similarities and thresholds

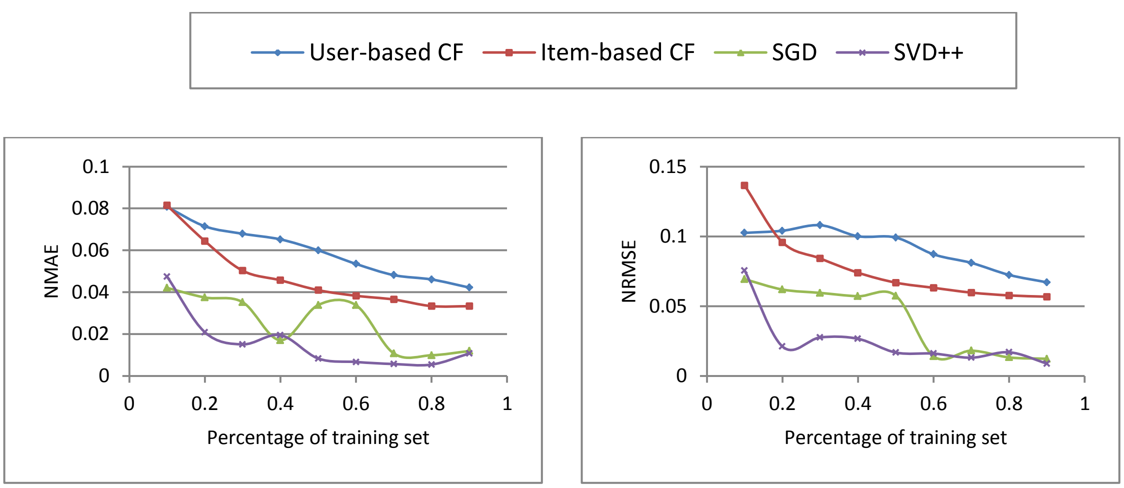
We measured NMAE and NRMSE at different neighborhood sizes as Figure 55.5 shows, and found that the Euclidean-weighting achieved the lowest NMAE and the lowest NRMSE.



(a) (b)

Figure 5.5 NMAE and NRMSE for user-based CF with different neighborhood sizes

We compared the performance of four algorithms that used PVR ratings at different percentages of the training set (Figure 5.6), and we found that SVD++ achieved the lowest NMAE and the lowest NRMSE.



(a) (b)

Figure 5.6 A performance comparison of different recommendation algorithms at different training ratios

We tested another PVR model, but no improvements were achieved during evaluation. For each user, we compared references added from a venue per year with the user's total added references from all venues, to determine the importance of that venue to the user in that particular year. However, this approach resulted in some issues; e.g., large venues were favored over small ones.

Although implicit rating is beneficial, some limitations exist. Users may add references to their libraries to be read at a later time, or they may never read articles they have added. Users may also choose to read an article based on its title, author, or abstract, none of which are directly related to the article's usefulness. Moreover, even if the researcher favors an article, this alone does not indicate the extent to which s/he favors it. Therefore, the articles in any given researcher's repository vary in terms of their importance to that researcher.

Using explicit data such as favorites or ratings for references could improve the accuracy of recommendations, as explicit data of this nature show that researchers are more or less interested in reading an article based on indications that they have read or liked it. 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. The explicit ratings can improve PVR measurements, especially in the case of researchers who have an interest in small-size venues. However, in order to collect data pertaining to these two fields, it would be necessary to construct a new dataset. The current dataset contains unique article IDs, rated only by the first researcher who added the article to CiteULike.

6. SUPPORTING THE CREATION OF SCHOLARLY BIBLIOGRAPHIES BY COMMUNITIES THROUGH ONLINE REPUTATION-BASED SOCIAL COLLABORATION*

6.1. Introduction

Closed bibliographic digital libraries (BDLs), whether manually compiled by authorized users or automatically generated, have existed for many years. In the last decade, open SRM websites (e.g., CiteULike and Bibsonomy[14]) have emerged. However, neither of these platforms achieves a level of precision or comprehensiveness sufficient to meet specific research needs. Current bibliographic search engines offer limited coverage of the available literature. No single search engine handles all the published articles in a subject area; thus, a search with any engine will return only a fraction of the available literature [301]. From this limited selection, researchers often concentrate further effort on specialized groups of publications, missing other valuable related research.

Many digital humanities projects manually maintain online BDLs that support diverse users in their efforts to locate a variety of references. In this section, our focal

---

* Reprinted with permission from "Supporting the Creation of Scholarly Bibliographies by Communities through Online Reputation Based Social Collaboration" by Hamed Alhoori, Omar Alvarez, Richard Furuta, Miguel Muñiz, Eduardo Urbina, 2009. Proceedings of the 13th European Conference on Research and Advanced Technology for Digital Libraries, Copyright 2009 Springer Berlin Heidelberg.

[14] http://www.bibsonomy.org/

example is the Cervantes[15] International Bibliography Online (CIBO), which aims to represent the best resources published since 1605 about Miguel de Cervantes, the author of *Don Quixote*. The resources are drawn from diverse multilingual and multicultural sources. The current CIBO bibliography gathering and filtering process is carried out by distinct sets of contributors: editors, reviewers, and authorized international collaborators. Consequently, delays of days or months can result before new publications are included in the CIBO, as the processes of gathering, filtering, and indexing must take place first.

The model followed by most online bibliographies is one wherein services are provided to users but users are not permitted to contribute to the bibliographies. This approach means that considerable external knowledge is not reflected in the results that bibliographies present to users. The current trend, supported by social platforms, however, is toward bilateral interaction, such that users can both benefit from the available knowledge and contribute to it. Hendry et al. [302] mention an "amateur bibliography" that is collected by non-professionals but falls short of the standards of a professional bibliography. This approach has the benefit of affording opportunities for a large number of references to be collected in a short span of time. However, there are also disadvantages inasmuch as this approach produces redundancy (duplicated references), spam, phantom author names, and phantom references. These do not support the level of high-quality scholarly research needed and expected from users [303]. Spam

[15] http://cervantes.tamu.edu/

also threatens social networking services by impairing contributions, interactions, and openness [304].

Social moderation models can be used to unify online groups and achieve consensus on topics of common interest, to reduce spam, and to provide information about members in regard to background and reputation. However, controversy exists pertaining to whether moderation in open environments is effective in producing content of an acceptable quality or if it is a reliable means of determining a user's reputation. Moderated systems face problems such as insufficient attention to posts on the part of moderators, moderation delays, unfair moderation decisions, and premature negative or positive consensus [305].

In this section, we propose an online reputation-based social collaboration (ORSC) approach to building a moderated scholarly bibliography [306] by benefiting from the "wisdom of the crowds" [307]. We experiment with this issue by implementing online social functionality for the CIBO using Drupal,[16] which is an open source content management platform. We test using a group of CIBO users to gather, evaluate, share, annotate, rank, and discover academic literature. We compared our precision outcomes with WorldCat,[17] which is a well-known union catalogue, the Modern Language Association International Bibliography (MLAIB[18]), and some SRM websites.

[16] https://www.drupal.org/
[17] http://www.worldcat.org
[18] http://www.mla.org/bibliography

6.2. Related Work

We compared the main features supported by various humanities BDLs as shown in Table 66.1. These BDLs are well established and most do not incorporate any social collaboration features such as social bookmarking, tagging, reviewing, or ranking.

Table 6.1 Humanities BDLs supported features

Features \ Bibliography	Cervantes Project	World Shakespeare Bibliography[19]	Galileo Project[20]	Walt Whitman Archive[21]
Developer	Texas A&M University	Shakespeare Quarterly	Rice University	University of Nebraska–Lincoln
Year established	1995	1960	1995	1995
Searching	√	√	√	√
Browsing	√	√	√	√
Multilanguage content	√	√	×	√
Multilanguage interface	√	×	×	×
Social collaboration	×	×	×	×

The ShaRef system [308] supports collaboration between groups of researchers and provides authentication and access control features. Heymann et al. [309] found that data provided by social bookmarking platforms could be unique, i.e., not available on any

19 http://www.worldshakesbib.org
20 http://galileo.rice.edu
21 http://www.whitmanarchive.org

other sources. Santos-Neto et al. [310] showed that very little collaboration takes place on CiteULike and Connotea. Online social platforms face spamming issues [304][311]. Bogers et al. [312] reported high spamming levels at BibSonomy and CiteULike. Krause et al. [311] mentioned that web spam has begun targeting scholarly communities and introduced some approaches to fight spam in social bookmarking services. We compared the main social collaboration features of four popular SRM websites, as shown in Table 66.2.

Table 6.2 Comparison of social features in SRM websites

Features \ **Social reference management**	**2collab**[22]	**BibSonomy**	**CiteUlike**	**Connotea**[23]
Multilanguage Interface	×	English and German	×	×
Social Bookmarking	√	√	√	√
Social Tagging	√	√	√	√
Social Reviewing	√	√	√	√
Social Ranking and Sorting	√	×	×	×
Social Filtering	√	√	×	×
Groups of Interest	√	√	√	√
Reputation-based Social Moderation	×	×	×	×

[22] http://www.2collab.com
[23] http://www.connotea.org

We found that all the websites included in the comparison support well-known social collaboration features and that each website used a distinct group type—private, closed, or open—to moderate references. In private groups, the community is hidden from nonmembers and only established members can contribute. In closed groups, moderators approve new members before the latter are permitted to contribute. In open groups, anyone can contribute; thus, there is an urgent need to check members' contributions. However, none of these types of group collaboration allow the community to collaborate fully.

All these groups assign moderators manually, which is time-consuming and may reflect some element of influence or bias on the part of the creators of the group. Further, no matter how they are assigned, moderators may lose interest or become inactive for long periods of time. Moreover, in the context of interdisciplinary bibliographies, determining if a reference is spam is likely to be challenging, as moderators may have insufficient knowledge of all related literature to support consistently correct judgments. To our knowledge, no bibliography is using or has attempted to use an approach such as ORSC.

6.3. Extending the CIBO to Support ORSC

6.3.1. Online Reputation-based Social Collaboration

Given the existence of spam on SRM websites, there is a need to reflect the quality of users' contributions to determine the reputation of any given user in a community in which users can play a the role of moderator. However, this need must be addressed in way that is balanced against the continuing need to benefit from the openness of social websites.

In addition to simply perusing a site and looking up a reference, users can participate in a bibliography site in a number of ways. They can add new references (C), tag existing references (T), rate references by assigning a score from 1 to 5 (R), review references by commenting on them (V), translate references (N), and filter spam references by marking them as such (F). In the present study, we designate three types of membership levels: user (u), collaborator, (b) and moderator (m). Users can search for and share references freely; however, their contributions are moderated. Contributions can be approved by a moderator or by n collaborators: $n=(1+ceiling(rc/ac))$ where rc and ac represent the rejected and approved contributions from collaborators, respectively. The higher the number of rejected contributions, the higher the number of collaborators needed to approve a new contribution.

Sabater and Sierra [313] present an extensive study of a set of reputation models in order to consider the nature of the social relationships among users. Chen et al. [314] present a user-reputation model used in a user-interactive question-and-answer system that combines social network analysis and user ratings. Jin et al. [315] present a user-reputation model for a digital library and digital education community that combines individual and collaborative activity.

Our model considers a user's activities and other users' evaluations of such activities. The elements selection and its assignment of weights are based on the experience of CIBO moderators. Members are upgraded or downgraded using a social reputation model [316], and they obtain a strong reputation (i.e., a high ranking) in the community by making accurate contributions and receiving credits from other members. A user can be upgraded to a collaborator, and a collaborator can be upgraded to a moderator. Initially, we seeded the moderator list with well-known Cervantes scholars and contributors. A summary of the contribution rules and privileges is shown in Table 6.3.

Table 6.3 Contribution rules

Controls / **Members**	**Create contribution**	**Approve contribution**	**Edit contribution**
User (u)	√	×	×
Collaborator (b)	√	√ nb	×
Moderator (m)	√	√	√

We summarize social reputation by using the following approach. If the summation of the user's (u) contributions S(u) and the summation of other users' evaluations of those contributions E(u), according to the importance of the contribution, the time of the contribution, and the evaluator's reputation (ER), exceeds a threshold value D, then the user (u) will be upgraded to a collaborator. If $(S(u)+E(u))>(D\times \log X)$, then the user will be upgraded to a moderator. X is the total number of contributions in the system.

S(u) is used to compute the user's contributions, as shown in Equation 6.1. *S(u)* is the total of the user's approved contributions of *C*, *T*, *R*, *V*, *N*, and *F* after these are multiplied by specified weights *a* to *f*, which represent the importance of that contribution. *X(u)* is the total of the user's approved contributions, where $X \in \{C, T, R, V, N, F\}$. X_i^u represents a single user (u) and that user's contributions (i). We also multiply the total of the user's contributions by the reciprocal of t_i and o_i, where t_i stands for the time that passed from the point at which the reference appeared in the literature to the point at which it was contributed to the CIBO, or the time from the contribution to the time of a follow-up contribution such as the addition of new tags, ratings, reviews, translations, or filters. o_i stands for the order of the contribution compared to other similar contributions that are related to a particular reference. This system allows users who at an earlier point have already made valid contributions related to a particular reference to gain more points that advance them to higher ranks in the community:

$$S(u) = a\sum_{i=1}^{C(u)}\left(\frac{C_i^u}{t_i}\right) + b\sum_{i=1}^{T(u)}\left(\frac{T_i^u}{ot_i}\right) + c\sum_{i=1}^{R(u)}\left(\frac{R_i^u}{ot_i}\right) + d\sum_{i=1}^{V(u)}\left(\frac{V_i^u}{ot_i}\right) + e\sum_{i=1}^{N(u)}\left(\frac{N_i^u}{ot_i}\right) + f\sum_{i=1}^{F(u)}\left(\frac{F_i^u}{ot_i}\right) \qquad (6.1)$$

To compute the users' evaluations, we use *E(u)* as shown in Equation 6.2. EX^u is a single evaluation of contribution X. *E(u)* provides the total of the users' evaluations (EX_{ij}^u) for a user's contributions after these are multiplied by a specified weight of *a'* to *e'* that again represents the importance of that contribution:

$$E(u) = a'\sum_{i=1}^{C_i^u}\sum_{j=1}^{EC_j^u}(EC_{ij}^u \times ER) + b'\sum_{i=1}^{T_i^u}\sum_{j=1}^{ET_j^u}(ET_{ij}^u \times ER) + c'\sum_{i=1}^{R_i^u}\sum_{j=1}^{ER_j^u}(ER_{ij}^u \times ER) +$$
$$d'\sum_{i=1}^{V_i^u}\sum_{j=1}^{EV_j^u}(EV_{ij}^u \times ER) + e'\sum_{i=1}^{N_i^u}\sum_{j=1}^{EN_j^u}(EN_{ij}^u \times ER) + f'\sum_{i=1}^{F_i^u}\sum_{j=1}^{EF_j^u}(EF_{ij}^u \times ER) \quad (6.2)$$

In order to compute *D*, we use Equation 6.3, where *U* stands for the total number of users, *J* the total number of rejected contributions, *A* the total number of approved contributions, and *E* the total number of evaluations:

$$D = \log(U) + \log(\frac{J}{A} \times E) \quad (6.3)$$

6.3.2. Social Technologies Applied to Bibliographies

A set of social collaboration features was implemented in CIBO to support an open social collaboration environment. Figure 6.1 shows the main interface displaying a reference's details.

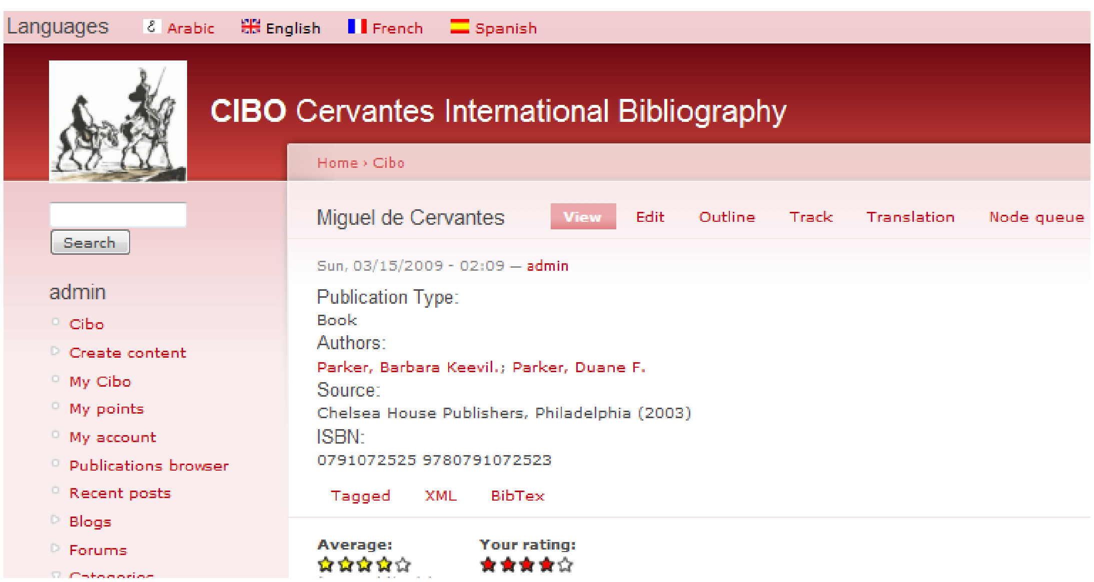


Figure 6.1 A screenshot of a reference's details

6.3.2.1. Social bookmarking

Users can participate by providing new references using the social bookmarking feature to import references or enter them manually. Figure 6.2 shows the points gained by a user after s/he has provided several contributions.

Points for admin

Points	Approved?	Date	Operation	Category	Description
1	Approved	04/27/2008 - 16:39	vote	Uncategorized	Vote cast: node 78.
1	Approved	04/24/2008 - 19:09	vote	Uncategorized	Vote cast: node 54.
1	Approved	04/24/2008 - 14:40	vote	Uncategorized	Vote cast: node 56.
1	Approved	04/24/2008 - 14:40	vote	Uncategorized	Vote cast: node 50.
2	Approved	04/24/2008 - 14:03	insert	Uncategorized	None
2	Approved	04/24/2008 - 14:02	insert	Uncategorized	None
-2	Approved	04/23/2008 - 00:45	operation	Uncategorized	None
1	Approved	04/22/2008 - 18:24	vote	Uncategorized	Vote cast: node 41.
2	Approved	04/22/2008 - 15:01	insert	Uncategorized	None
1	Approved	04/22/2008 - 14:58	vote	Uncategorized	Vote cast: node 40.

Uncategorized points Balance: 1503

Approved points Balance: 1503
Points awaiting moderation: 0
Net points Balance: 1503

« first ‹ previous … 72 73 74 75 76 77 78 **79** 80 next › last »

Figure 6.2 Detailed view of a user's points

6.3.2.2. Social tagging

Delicious[24] and Digg[25] are popular social web services that use folksonomy tagging. In open environments such as these, misleading and inaccurate tags are common—even expected. However, this is not acceptable in scholarly research communities. In CIBO, our goal is to prevent these effects by moderating new users' tags. Users can create their own tags or reuse previously approved tags.

[24] https://delicious.com
[25] http://digg.com

6.3.2.3. Social ranking

Bibliography ranking has been used as a way to give users top-N resources from the search results. In the present study, users rated references on a scale of 1 to 5 points.

6.3.2.4. Social reviewing

We implemented a feedback environment in order to build an active online research community. The environment provides a place where users can post and read reviews and comments.

6.3.2.5. Social translation

As digital libraries expand their audience and content scope, there is an increasing need for resources and access tools for those resources in a variety of languages [317]. The Cervantes Project's international scope requires the inclusion of content and system functionalities in multiple languages, as Cervantes literature has been translated into various languages and a goal of the CIBO project is to establish bridges between cultures.

Users can choose the language they prefer to use from those available in a the system. The interface display is then automatically translated into the language chosen by the user, and following the system selects only content in that language. Using the Google Translate API,[26] we provided a translation capability for the comments. Bibliographic data can be entered in a language and then manually translated into another language and/or linked to existing bibliographic data or publications in other languages (Figure 6.3).

[26] http://code.google.com/apis/ajaxlanguage

Figure 6.3 Available translations of a publication

6.3.2.6. Social filtering

Retrieving references that are irrelevant, incorrect, or spam frustrates researchers and has a negative impact on their productivity. We tried to address this scenario to some extent by empowering users to discover and filter results of this nature and to report such results and spammers for moderation. A moderator or *n* collaborators can approve requests by editing or hiding contributions or by banning a spammer. Moderators are able to view these changes for any follow-up requests.

6.4. Experiments

We used WorldCat, the MLAIB, and four SRM systems (CiteULike, Connotea, 2collab, and BibSonomy), which together comprise millions of references. We compared the precision outcomes of each of these bibliographies and SRM systems with those of the augmented CIBO. Precision in our experiments was calculated as the number of relevant references retrieved by a search divided by the total number of references retrieved by that search at several milestones. Cervantes project contributors determined the most common keywords and tags used in Cervantes literature, which we used as search terms. These contributors also determined the relevance of the retrieved references. After gathering the results from the different resources, we found that Connotea and 2collab contain only a few references about Cervantes. Therefore, we removed them from the comparison. Table 66.4 shows a sample of precision for the first 10 retrieved references as compared across CiteULike, BibSonomy, WorldCat, and MLAIB. We used keywords and tags in combinations of various lengths to search the bibliographies.

Table 6.4 A sample of P@10 on various platforms

BDLs / **Search terms**	**WorldCat**	**MLAIB**	**CiteULike**	**BibSonomy**	**CIBO**
Cervantes	80	100	30	30	100
سيرفانتس	0	0	0	0	40
Quixote	100	90	50	50	90
Quijote	100	90	50	50	90
Cervantes plays	90	40	30	00	80
Miguel de Cervantes Poetry	30	10	0	0	100
Cervantes Windmills	80	100	30	10	80
Sancho Panza	100	100	20	0	100
Dulcinea	80	80	10	0	50
Cervantes Blanket	10	30	10	0	0
Cervantes Island	30	30	0.0	0	90
Cervantes Persiles	80	70	10	0	90

Figure 6.46 shows the average precision percentage at 10 (P@10), 20, 30, 40, and 50. The figure shows that CIBO performs better than the other websites at precision 10.

At precision 20, CIBO is still ahead of WorldCat by 2%. At precision 30, however, WorldCat moves ahead by 1%. This pattern occurred mainly because users rated and filtered the initial results but neglected the subsequent outcomes.

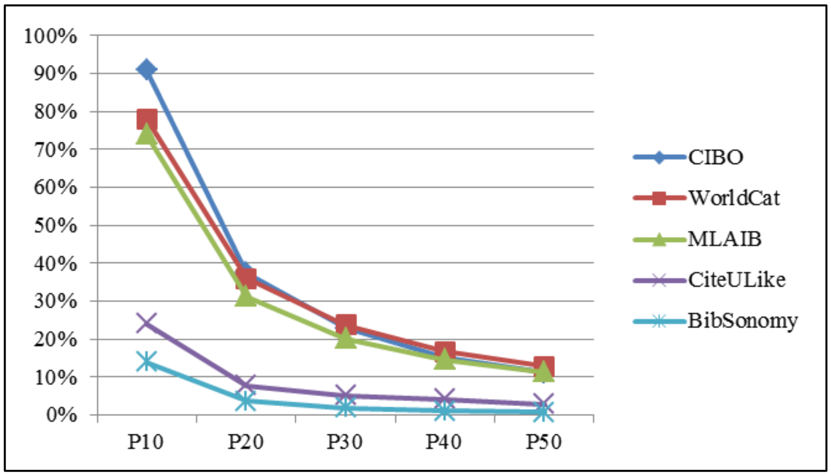


Figure 6.4 Precision of the compared BDLs with CIBO

Our findings show that the closed BDLs studied are considerably more precise in regard to the results they return than SRM systems are. This finding supports the argument that scholarly communities should continue to maintain closed environments but this would also increase the limited scope of coverage on literature. However, on searches for Cervantes-related topics, at least, the ORSC approach produces better precision outcomes than closed bibliographies do.

7. CONCLUSION AND FUTURE WORK

This section presents a summary of this dissertation, its contributions to the field and plans to extend this research.

7.1 Summary

Given the proliferation of scholarly products, it is becoming challenging for researchers to remain up-to-date with new findings. Additionally, increasingly complex multidisciplinary research areas are emerging and researchers' interests are shifting over time. Previous studies have focused on using citation analysis, an approach that demonstrated several drawbacks.

This dissertation explored the influence of the social web on scholarly communities and investigated several methods and techniques to utilize web-based indicators to support such communities and reduce information overload. First, this dissertation studied international scholarly information behavior and addressed several scholarly needs and expectations. Second, it studied non-traditional web-based indicators at various levels and used them to predict scholarly and social impact. Third, it utilized such web indicators to recommend scholarly venues related to any given researcher's readings and interests. Finally, we developed a scholarly bibliography using reputation-based social collaboration, which is considerably more comprehensive and accurate than other bibliographies.

7.2 Contributions and Plans

Section 2 investigated current practices and scholarly activities on an international level in the social media age. We compared the scholarly information behavior and information needs of researchers in the United States and Qatar. The survey revealed several significant relationships that deepen our overall understanding of scholarly attitudes. For example, we found that 40% of SRM users search for articles within SRMs, and that SRM users use more tags and are able to retrieve more articles related to their research. We found a number of similarities among the behaviors and needs of researchers in both studies. We also found that SRMs play an important role for students in finding and organizing scholarly articles and connecting with other researchers.

The study showed that publication overload continues to affect researchers. The researchers who had built a personal article collection were more satisfied with their information needs than others who did not have a collection of this nature. We found that scholarly information sources and tools are not being fully utilized. Moreover, even with all the advances in scholarly and social platforms, researchers' information needs are not yet being fully met.

Current academic digital libraries and SRMs are based on a "one size fits all" approach, but newer implementations should seek to address the specific needs of different disciplines and researchers. Many researchers become comfortable with the tools they are using such that new technologies must come with very clear benefits if researchers are to become motivated to try them.

In the future, a quantitative study is planned on a wider group of researchers and will investigate the specific research needs of different disciplines. George et al. [117] found that nearly all graduate students (96%) reported that academics influence their research and information seeking. We would like to investigate whether SRMs have any significant effect on research groups in building online collaborative research communities. Collaborative and social information seeking [318][319][320] has been studied and modeled to understand group work and activities. We intend to investigate the effects of SRMs on the research process and develop a collaborative research model of dynamic strategies. We will investigate scholarly information behavior among researchers producing or dealing with non-English content. Additionally, we plan to investigate how social media can build and affect a research culture.

In Section 3, altmetrics were explored at four different levels. We proposed and investigated JSI, our new measure computed using non-citation-based metrics, and compared it with several citation-based metrics. Significant correlations were found between JSI and IF, 5-IF, Immediacy Index, SJR, and article influence score. These findings suggest that, at least for the time being, journal rankings remain a trusted proxy for the quality of scholarly social media attention. Although altmetrics have the potential to predict delayed citation-based metrics, the latter metrics can also be used to validate the former. We also found that usage and coverage of social media for research activities is high on a few platforms.

JSI will be compared with itself as well as with citation-based metrics over a number of years in order to check the validity and reliability of altmetrics. A theoretical multi-dimensional model will be built to improve the overall understanding of altmetrics. Other factors that may influence altmetrics will be investigated such as publishers, disciplines, journal age, submission and acceptance rates, and reputation of editorial board members. Also slated for further scrutiny are the effects of features such as article details. Further, plans include the close examination of scholarly mentions in online news from different angles, such as size and geographic location (e.g., local, national, and international).

The study will be extended to encompass more countries and to explore whether altmetrics can be used as a basis for determining the local social impact of research and emerging research interests across nations. The investigation will look at why the altmetrics coverage was particularly high for some countries and how altmetrics can be used when major social media tools are blocked in other countries.

The study also explored the relationship between altmetrics and NOA and OA articles. On eight online platforms (F1000, Facebook, CiteULike, Mendeley, peer review sites, Twitter, Reddit, and blogs), the results showed that OA articles received more altmetrics than NOA articles. However, when investigating the effects of some influential factors such as journal, publication year, and citation count, less significant differences between OA and altmetrics were found. We found that academic social networks had a high OAAA. However, the general social media sites differed in terms of the quantity of altmetrics received between NOA and OA articles. For example, Facebook had a high OAAA, whereas Weibo had no OAAA. This study also reported a significant correlation between citations and altmetrics for NOA and OA articles, which was not the case in some previous studies that compared articles in general [11]. Plans are in place to expand this part of the study to include more journals and articles and to explore disciplinary differences, as well as to investigate whether and to what extent there are differences in altmetrics between green and gold OA articles [5].

In Section 4, a new multi-dimensional approach was described that can measure, in real-time, the impact of digital humanities research using academic social media sites. The findings indicate that RCAR and altmetrics can quantify an early scholarly impact of articles. Also investigated was the relationship between ranking methods for scholarly venues that use traditional citation-based metrics and our proposed social-based metrics. Statistically significant correlations were found between the two approaches, with disciplinary differences. The findings suggest that SRM systems have the potential to provide an early intellectual indicator of the influence of scholarly venues and to reduce the limitations of citation-based metrics.

In the future, more studies will be conducted to better understand how these observations reflect the needs and standards of a given field. I plan to investigate whether a single set of social-based metrics can effectively measure the influence of scholarly venues in all research areas, or whether it is necessary for each research area to define its own metrics. I also plan to explore how data from SRM systems differ and whether they measure similar or different impacts of research. The PLOS has announced that its articles have received more than 500 million altmetrics events.[27] This figure suggests that, with millions of articles already published, billions of altmetrics events are

[27] http://blogs.plos.org/tech/lessons-learned-developing-scholarly-open-source-software/

waiting to be analyzed and modeled. I intend to build multidimensional models to evaluate and predict trusted social, cultural, environmental, and economic research impacts. These models will make sense of new complex and large distributed datasets and enrich our understanding and usage of scholarly outcomes.

Multidisciplinary research areas are rapidly emerging, and the number of scholarly venues is growing. Researchers need to discover venues of interest to them, and research institutions need to be aware of these venues. In Section 5, using data from an academic social network, an approach is described to recommend potential scholarly venues for researchers to follow or to publish in based on their current interests.

A new weighting strategy was developed for rating venues based not only on personal references but also on the temporal factor of when the references were added. Experiments with this strategy in the recommendation process using a real data set produced results that showed improvements in accuracy and ranking quality compared with a 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 from a venue, or in general, 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 my future research, I plan to enhance the recommendation quality by using measures such as a researcher's trustworthiness and reputation [15] with the goal of improving accuracy, diversity, novelty, and serendipity [321]. Also planned is a user study through which I will collect explicit ratings to compare with our implicit ratings. These results will be used to recommend venues for manuscripts. Along these same lines, the system will begin using metadata of articles, such as title, abstract, keywords, and tags, to recommend venues. These experiments will use a hybrid approach implementing both CF and content-based filtering. In addition, other factors will be considered, such as budget availability and the ability to travel in cases such as conferences or workshops.

Open bibliography environments were originally conceived as websites for exchanging references and reviews of global publications within large communities on the Internet. These sites offer a variety of benefits, but the lack of moderation means that the results they return are not as relevant as those returned by bibliography environments that do include moderation. A lack of moderation may be acceptable for social websites but is inappropriate in scholarly communities, where content quality is a priority.

In Section 6, the investigation examined the precision outcomes of a hybrid bibliography system created by an online digital humanities community. Experimental results indicate that ORSC improves the quality and credibility of SRM websites. In the future, additional automation of the moderation process will compare the contributed references in the system discussed in this section against the references retrieved from closed and open social reference management websites.

REFERENCES

[1] Cambridge Economic Policy Associates, “Activities, costs and funding flows in the scholarly communications system,” *Research Information Network*, 2008. [Online]. Available: http://www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf.

[2] M. M. Cummings, “Publications: Progress or pollution,” *Am. Sci.*, vol. 61, no. 2, pp. 163–166, Mar. 1973.

[3] C. Tenopir, D. W. King, S. Edwards, and L. WU, “Electronic journals and changes in scholarly article seeking and reading patterns,” *Aslib Proc.*, vol. 61, no. 1, pp. 5–32, Jan. 2009.

[4] Z. Liu, “Reading behavior in the digital environment: Changes in reading behavior over the past ten years,” *J. Doc.*, vol. 61, no. 6, pp. 700–712, Dec. 2005.

[5] S. Harnad, T. Brody, F. Vallières, L. Carr, S. Hitchcock, Y. Gingras, C. Oppenheim, C. Hajjem, and E. R. Hilf, “The access/impact problem and the green and gold roads to open access: An update,” *Ser. Rev.*, vol. 34, no. 1, pp. 36–40, Dec. 2008.

[6] M. Khabsa and C. L. Giles, “The number of scholarly documents on the public web,” *PLoS ONE*, vol. 9, no. 5, p. e93949, Jan. 2014.

[7] D. N. Boote and P. Beile, “Scholars before researchers: On the centrality of the dissertation literature review in research preparation,” *Educ. Res.*, vol. 34, no. 6, pp. 3–15, Aug. 2005.

[8] P. U. Kuruppu and A. M. Gruber, “Understanding the information needs of academic scholars in agricultural and biological sciences,” *J. Acad. Librariansh.*, vol. 32, no. 6, pp. 609–623, Nov. 2006.

[9] J. Murphy, “Information-seeking habits of environmental scientists: A study of interdisciplinary scientists at the environmental protection agency in research triangle park, North Carolina,” *Issues Sci. Technol. Librariansh.*, vol. 38, 2003.

[10] H. Alhoori and R. Furuta, “Do altmetrics follow the crowd or does the crowd follow altmetrics?,” in *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2014, pp. 375–378.

[11] H. Alhoori, R. Furuta, M. Tabet, M. Samaka, and E. A. Fox, “Altmetrics for country-level research assessment,” in *Proceedings of the 16th International Conference on Asia-Pacific Digital Libraries*, 2014, vol. 8839, pp. 59–64.

[12] H. Alhoori, S. R. Choudhury, T. Kanan, R. Furuta, E. A. Fox, and C. L. Giles, “On the relationship between open access and altmetrics,” in *Proceedings of the iConference*, 2015.

[13] H. Alhoori and R. Furuta, “Identifying the real-time impact of the digital humanities using social media measures,” in *Proceedings of the Digital Humanities Conference*, 2013.

[14] H. Alhoori and R. Furuta, “Can social reference management systems predict a ranking of scholarly venues?,” in *Proceedings of the 17th international conference on theory and practice of digital libraries*, 2013, vol. 8092, pp. 138–143.

[15] H. Alhoori, O. Alvarez, R. Furuta, N. Miguel Mu, and E. Urbina, “Supporting the creation of scholarly bibliographies by communities through online reputation based social collaboration,” in *Proceedings of the 13th European conference on research and advanced technology for digital libraries*, 2009, vol. 5714, pp. 180–191.

[16] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr, “Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 12, pp. 2351–2369, Dec. 2012.

[17] C. L. Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press, 2007.

[18] T. Kortelainen and M. Katvala, “‘Everything is plentiful—Except attention’. Attention data of scientific journals on social web tools,” *J. Informetr.*, vol. 6, no. 4, pp. 661–668, Oct. 2012.

[19] U. Farooq, Y. Song, J. M. Carroll, and C. L. Giles, “Social bookmarking for scholarly digital libraries,” *IEEE Internet Comput.*, vol. 11, no. 6, pp. 29–35, Nov. 2007.

[20] T. Bogers and A. Van Den Bosch, “Recommending scientific articles using citeulike,” in *Proceedings of the 2008 ACM conference on Recommender systems*, 2008, pp. 287–290.

[21] K. Emamy and R. Cameron, “CiteULike: A researcher’s social bookmarking service,” *Ariadne*, vol. 51, Apr. 2007.

[22] T. E. Vanhecke, “Zotero,” *J. Med. Libr. Assoc.*, vol. 96, no. 3, pp. 275–276, Jul. 2008.

[23] D. Benz, A. Hotho, R. Jäschke, B. Krause, F. Mitzlaff, C. Schmitz, and G. Stumme, “The social bookmark and publication management system bibsonomy,” *Int. J. Very Large Data Bases*, vol. 19, no. 6, pp. 849–875, Dec. 2010.

[24] V. Henning and J. Reichelt, “Mendeley - a last.fm for research?,” in *IEEE Fourth International Conference on eScience*, 2008, pp. 327–328.

[25] M. Thelwall and K. Kousha, “Academia.edu: Social network or academic network?,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 721–731, Apr. 2014.

[26] M. Thelwall and K. Kousha, “ResearchGate: Disseminating, communicating, and measuring scholarship?,” *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 5, pp. 876–889, May 2015.

[27] S. Makri and C. Warwick, “Information for inspiration: Understanding architects’ information seeking and use behaviors to inform design,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 9, pp. 1745–1770, Sep. 2010.

[28] H. K. Sahu and S. Nath Singh, “Information seeking behaviour of astronomy/astrophysics scientists,” *Aslib Proc.*, vol. 65, no. 2, pp. 109–142, Feb. 2013.

[29] C. Tenopir, D. W. King, P. Boyce, M. Grayson, and K.-L. Paulson, “Relying on electronic journals: Reading patterns of astronomers,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 56, no. 8, pp. 786–802, Jun. 2005.

[30] J. Hoppenfeld and M. M. Smith, “Information-seeking behaviors of business faculty,” *J. Bus. Financ. Librariansh.*, vol. 19, no. 1, pp. 1–14, Jan. 2014.

[31] D. Flaxbart, “Conversations with chemists,” *Sci. Technol. Libr.*, vol. 21, no. 3–4, pp. 5–26, Nov. 2001.

[32] P. M. Davis, “Information-seeking behavior of chemists: A transaction log analysis of referral URLs,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 55, no. 4, pp. 326–332, Feb. 2004.

[33] K. Athukorala, E. Hoggan, A. Lehtiö, T. Ruotsalo, and G. Jacucci, “Information-seeking behaviors of computer scientists: Challenges for electronic literature

search tools," in *Proceedings of the American Society for Information Science and Technology*, 2013, vol. 50, no. 1, pp. 1–11.

[34] J. Bichteler and D. Ward, "Information-seeking behaviour of geoscientists," *Spec. Libr.*, vol. 80, no. 3, pp. 169–178, 1989.

[35] S. E. Wiberley and W. G. Jones, "Patterns of information seeking in the humanities," *Coll. Res. Libr.*, vol. 50, no. 6, pp. 638–645, Nov. 1989.

[36] A. Barrett, "The information-seeking habits of graduate student researchers in the humanities," *J. Acad. Librariansh.*, vol. 31, no. 4, pp. 324–331, Jul. 2005.

[37] G. Buchanan, S. J. Cunningham, A. Blandford, J. Rimmer, and C. Warwick, "Information seeking by humanities scholars," in *Proceedings of the 9th European Conference on Digital Libraries*, 2005, vol. 3652, pp. 218–229.

[38] M. L. Cohen, "Research habits of lawyers," *Jurimetrics J.*, vol. 9, no. 4, pp. 183–194, Jun. 1969.

[39] Stephann Makri, Ann Blandforda, and Anna L. Coxa, "Investigating the information-seeking behaviour of academic lawyers: From Ellis's model to design," *Inf. Process. Manag.*, vol. 44, no. 2, pp. 613–634, Mar. 2008.

[40] M. A. Wilkinson, "Information sources used by lawyers in problem solving: An empirical exploration," *Library & Information Science Research*, vol. 23, no. 3. pp. 257–276, Sep-2001.

[41] J. Zhao, M.-Y. Kan, and Y. L. Theng, "Math information retrieval: User requirements and prototype implementation," in *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2008, pp. 187–196.

[42] K. Davies, "The information-seeking behaviour of doctors: a review of the evidence," *Health Info. Libr. J.*, vol. 24, no. 2, pp. 78–94, Jun. 2007.

[43] K. L. Curtis, A. C. Weller, and J. M. Hurd, "Information-seeking behavior of health sciences faculty: The impact of new information technologies," *Bull. Med. Libr. Assoc.*, vol. 85, no. 4, pp. 402–410, Oct. 1997.

[44] C. Dee and E. E. Stanley, "Information-seeking behavior of nursing students and clinical nurses: Implications for health sciences librarians," *J. Med. Libr. Assoc.*, vol. 93, no. 2, pp. 213–222, Apr. 2005.

[45] D. Revere, A. M. Turner, A. Madhavan, N. Rambo, P. F. Bugni, A. Kimball, and S. S. Fuller, "Understanding the information needs of public health practitioners:

A literature review to inform design of an interactive digital knowledge management system," *J. Biomed. Inform.*, vol. 40, no. 4, pp. 410–421, Aug. 2007.

[46] N. L. Pelzer, W. H. Wiese, and J. M. Leysen, "Library use and information-seeking behavior of veterinary medical students revisited in the electronic environment," *Bull. Med. Libr. Assoc.*, vol. 86, no. 3, pp. 346–355, Jul. 1998.

[47] I. Rowlands, D. Nicholas, P. Williams, P. Huntington, M. Fieldhouse, B. Gunter, R. Withey, H. Jamali, T. Dobrowolski, and C. Tenopir, "The Google generation: The information behaviour of the researcher of the future," *Aslib Proc.*, vol. 60, no. 4, pp. 290–310, Jun. 2008.

[48] C. Warwick, J. Rimmer, A. Blandford, J. Gow, and G. Buchanan, "Cognitive economy and satisficing in information seeking: A longitudinal study of undergraduate information behavior," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 12, pp. 2402–2415, Dec. 2009.

[49] E. Whitmire, "Disciplinary differences and undergraduates' information-seeking behavior," *J. Am. Soc. Inf. Sci. Technol.*, vol. 53, no. 8, pp. 631–638, Jan. 2002.

[50] C. M. Brown, "Information literacy of physical science graduate students in the information age," *Coll. Res. Libr.*, vol. 60, no. 5, pp. 426–438, Sep. 1999.

[51] A. Catalano, "Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature," *J. Doc.*, vol. 69, no. 2, pp. 243–274, 2013.

[52] C. Tenopir, D. W. King, P. Boyce, M. Grayson, Y. Zhang, and M. Ebuen, "Patterns of journal use by scientists through three evolutionary phases," *D-Lib Mag.*, vol. 9, no. 5, May 2003.

[53] H. R. Jamali and S. Asadi, "Google and the scholar: The role of Google in scientists' information-seeking behaviour," *Online Inf. Rev.*, vol. 34, no. 2, pp. 282–294, Apr. 2010.

[54] M. Hertzum and A. M. Pejtersen, "The information-seeking practices of engineers: Searching for documents as well as for people," *Inf. Process. Manag.*, vol. 36, no. 5, pp. 761–778, Sep. 2000.

[55] D. Engel, S. Robbins, and C. Kulp, "The information-seeking habits of engineering faculty," *Coll. Res. Libr.*, vol. 72, no. 6, pp. 548–567, Nov. 2011.

[56] S. Robbins, D. Engel, and C. Kulp, "How unique are our users? Comparing responses regarding the information-seeking habits of engineering faculty," *Coll. Res. Libr.*, vol. 72, no. 6, pp. 515–532, Nov. 2011.

[57] B. M. Hemminger, D. Lu, K. T. L. Vaughan, and S. J. Adams, “Information seeking behavior of academic scientists,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 14, pp. 2205–2225, Dec. 2007.

[58] D. W. King, C. Tenopir, C. H. Montgomery, and S. E. Aerni, “Patterns of journal use by faculty at three diverse universities,” *D-Lib Mag.*, vol. 9, no. 10, Oct. 2003.

[59] S. Hirsh and J. Dinkelacker, “Seeking information in order to produce information: An empirical study at Hewlett Packard Labs,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 55, no. 9, pp. 807–817, Jul. 2004.

[60] D. Rusch-Feja and U. Siebeky, “Evaluation of usage and acceptance of electronic journals,” *D-Lib Mag.*, vol. 5, no. 10, Oct. 1999.

[61] E. T. Smith, “Changes in faculty reading behaviors: The impact of electronic journals on the University of Georgia,” *J. Acad. Librariansh.*, vol. 29, no. 3, pp. 162–168, May 2003.

[62] B. M. Hemminger and J. TerMaat, “Annotating for the world : Attitudes toward sharing scholarly annotations,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 11, pp. 2278–2292, Nov. 2014.

[63] S. Talja and H. Maula, “Reasons for the use and non-use of electronic journals and databases: A domain analytic study in four scholarly disciplines,” *J. Doc.*, vol. 59, no. 6, pp. 673–691, Dec. 2003.

[64] L. Haglund and P. Olsson, “The impact on university libraries of changes in information behavior among academic researchers: a multiple case study,” *J. Acad. Librariansh.*, vol. 34, no. 1, pp. 52–59, Jan. 2008.

[65] M. Macedo-Rouet, J.-F. Rouet, C. Ros, and N. Vibert, “How do scientists select articles in the PubMed database? An empirical study of criteria and strategies,” *Rev. Eur. Psychol. Appliquée/European Rev. Appl. Psychol.*, vol. 62, no. 2, pp. 63–72, Apr. 2012.

[66] C. Brown, “Where do molecular biology graduate students find information?,” *Sci. Technol. Libr.*, vol. 25, no. 3, pp. 89–104, Apr. 2005.

[67] E. T. Smith, “Assessing collection usefulness: An investigation of library ownership of the Resources Graduate Students Use,” *Coll. Res. Libr.*, vol. 64, no. 5, pp. 344–355, Sep. 2003.

[68] J. M. Hurd, D. D. Blecic, and R. Vishwanatham, "Information use by molecular biologists: Implications for library collections and services," *Coll. Res. Libr.*, vol. 60, no. 1, pp. 31–43, Jan. 1999.

[69] L. Zhang, "Discovering information use in agricultural economics: A citation study," *J. Acad. Librariansh.*, vol. 33, no. 3, pp. 403–413, May 2007.

[70] P. Kelsey and T. Diamond, "Establishing a core list of journals for forestry: A citation analysis from faculty at southern universities," *Coll. Res. Libr.*, vol. 64, no. 5, pp. 357–377, Sep. 2003.

[71] M. A. Burright, T. B. Hahn, and M. J. Antonisse, "Understanding information use in a multidisciplinary field: A local citation analysis of neuroscience research," *Coll. &Research Libr.*, vol. 66, no. 3, pp. 198–211, May 2005.

[72] D. Bolchini, A. Finkelstein, V. Perrone, and S. Nagl, "Better bioinformatics through usability analysis," *Bioinformatics*, vol. 25, no. 3, pp. 406–412, Feb. 2009.

[73] P. M. Davis and L. R. Solla, "An IP-level analysis of usage statistics for electronic journals in chemistry: Making inferences about user behavior," *J. Am. Soc. Inf. Sci. Technol.*, vol. 54, no. 11, pp. 1062–1068, Sep. 2003.

[74] D. Nicholas, P. Huntington, H. R. Jamali, and A. Watkinson, "The information seeking behaviour of the users of digital scholarly journals," *Inf. Process. Manag.*, vol. 42, no. 5, pp. 1345–1365, Sep. 2006.

[75] H. R. Ke, R. Kwakkelaar, Y. M. Tai, and L. C. Chen, "Exploring behavior of e-journal users in science and technology: Transaction log analysis of Elsevier's ScienceDirect onsite in Taiwan," *Libr. Inf. Sci. Res.*, vol. 24, no. 3, pp. 265–291, Sep. 2002.

[76] A. Rozic-Hristovsk, D. Hristovski, and L. Todorovski, "Users' information-seeking behavior on a medical library website," *J. Med. Libr. Assoc.*, vol. 90, no. 2, pp. 210–217, Apr. 2002.

[77] D. D. Blecic, J. L. Dorsch, M. H. Koenig, and N. S. Bangalore, "A longitudinal study of the effects of OPAC screen changes on searching behavior and searcher success," *Coll. Res. Libr.*, vol. 60, no. 6, pp. 515–530, Nov. 1999.

[78] S. Jones, S. J. Cunningham, R. McNab, and S. Boddie, "A transaction log analysis of a digital library," *Int. J. Digit. Libr.*, vol. 3, no. 2, pp. 152–169, Aug. 2000.

[79] R. Islamaj Dogan, G. C. Murray, A. Névéol, and Z. Lu, "Understanding PubMed® user search behavior through log analysis," *Database (Oxford).*, vol. 2009, Nov. 2009.

[80] D. Nicholas, P. Huntington, and H. R. Jamali, "Diversity in the information seeking behaviour of the virtual scholar: Institutional comparisons," *J. Acad. Librariansh.*, vol. 33, no. 6, pp. 629–638, Dec. 2007.

[81] G. Marchionini, *Information Seeking in Electronic Environments*. NY, USA: Cambridge University Press, 1995.

[82] T. D. Wilson, "Models in information behaviour research," *J. Doc.*, vol. 55, no. 3, pp. 249–270, Aug. 1999.

[83] L. I. Meho and H. R. Tibbo, "Modeling the information-seeking behavior of social scientists: Ellis's study revisited," *J. Am. Soc. Inf. Sci. Technol.*, vol. 54, no. 6, pp. 570–587, Apr. 2003.

[84] M. A. Jankowska, "Identifying university professors' information needs in the challenging environment of information and communication technologies," *J. Acad. Librariansh.*, vol. 30, no. 1, pp. 51–66, Jan. 2004.

[85] B. T. Fidzani, "Information needs and information-seeking behaviour of graduate students at the University of Botswana," *Libr. Rev.*, vol. 47, no. 7, pp. 329–340, Nov. 1998.

[86] J. A. Adams and S. C. Bonk, "Electronic information technologies and resources: Use by university faculty and faculty preferences for related library services," *Coll. Res. Libr.*, vol. 56, no. 2, pp. 119–131, Mar. 1995.

[87] H. G. Rempel, "A longitudinal assessment of graduate student research behavior and the impact of attending a library literature review workshop," *Coll. Res. Libr.*, vol. 71, no. 6, pp. 532–547, Nov. 2010.

[88] H. G. Rempel and J. Davidson, "Providing information literacy instruction to graduate students through literature review workshops," *Issues Sci. Technol. Librariansh.*, vol. 53, 2008.

[89] M. Harrison, S. Summerton, and K. Peters, "EndNote training for academic staff and students: the experience of the Manchester Metropolitan University Library," *New Rev. Acad. Librariansh.*, vol. 11, no. 1, pp. 31–40, Apr. 2005.

[90] X. Niu, B. M. Hemminger, C. Lown, S. Adams, C. Brown, A. Level, M. McLure, A. Powers, M. R. Tennant, and T. Cataldo, "National study of information

seeking behavior of academic researchers in the United States," *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 5, pp. 869–890, May 2010.

[91] X. Niu and B. M. Hemminger, "A study of factors that affect the information-seeking behavior of academic scientists," *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 2, pp. 336–353, Feb. 2012.

[92] V. Larivière, C. R. Sugimoto, and P. Bergeron, "In their own image? A comparison of doctoral students' and faculty members' referencing behavior," *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. 5, pp. 1045–1054, May 2013.

[93] J. Bar-Ilan, "Information hub blogs," *J. Inf. Sci.*, vol. 31, no. 4, pp. 297–307, Aug. 2005.

[94] L. Bonetta, "Scientists enter the blogosphere," *Cell*, vol. 129, no. 3, pp. 443–445, May 2007.

[95] S. Kjellberg, "I am a blogging researcher: Motivations for blogging in a scholarly context," *First Monday*, vol. 15, no. 8, Aug. 2010.

[96] G. Kirkup, "Academic blogging: academic practice and academic identity," *London Rev. Educ.*, vol. 8, no. 1, pp. 75–84, Mar. 2010.

[97] J. Priem and K. L. Costello, "How and why scholars cite on Twitter," in *Proceedings of the American Society for Information Science and Technology*, 2010, vol. 47, no. 1, pp. 1–4.

[98] J. Letierce, A. Passant, S. Decker, and J. G. Breslin, "Understanding how Twitter is used to spread scientific messages," in *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.

[99] D. Ponte and J. Simon, "Scholarly communication 2.0: Exploring researchers' opinions on web 2.0 for scientific knowledge creation, evaluation and dissemination," *Ser. Rev.*, vol. 37, no. 3, pp. 149–156, Sep. 2011.

[100] F. Gu and G. Widén-Wulff, "Scholarly communication and possible changes in the context of social media: A Finnish case study," *Electron. Libr.*, vol. 29, no. 6, pp. 762–776, Sep. 2011.

[101] I. Rowlands, D. Nicholas, B. Russell, N. Canty, and A. Watkinson, "Social media use in the research workflow," *Learn. Publ.*, vol. 24, no. 3, pp. 183–195, Jul. 2011.

[102] R. Procter, R. Williams, J. Stewart, M. Poschen, H. Snee, A. Voss, and M. Asgari-Targhi, "Adoption and use of Web 2.0 in scholarly communications," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 368, no. 1926, pp. 4039–4056, Sep. 2010.

[103] K.-S. Kim, S.-C. J. Sin, and T.-I. Tsai, "Individual differences in social media use for information seeking," *J. Acad. Librariansh.*, vol. 40, no. 2, pp. 171–178, Mar. 2014.

[104] H. S. Du, S. K. W. Chu, G. E. Gorman, and F. L. C. Siu, "Academic social bookmarking: An empirical analysis of Connotea users," *Libr. Inf. Sci. Res.*, vol. 36, no. 1, pp. 49–58, Jan. 2014.

[105] A. Borrego and J. Fry, "Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy," *J. Inf. Sci.*, vol. 38, no. 3, pp. 297–308, Apr. 2012.

[106] E. Francese, "Usage of reference management software at the University of Torino," *JLIS.it*, vol. 4, no. 2, pp. 145–174, Jul. 2013.

[107] C. Tenopir, R. Volentine, and D. W. King, "Social media and scholarly reading," *Online Inf. Rev.*, vol. 37, no. 2, pp. 193–216, Apr. 2013.

[108] A. Gruzd and M. Goertzen, "Wired academia: Why social science scholars are using social media," in *Proceedings of the 46th Hawaii International Conference on System Sciences*, 2013, pp. 3332–3341.

[109] A. Mandavilli, "Peer review: Trial by Twitter," *Nature*, vol. 469, no. 7330, pp. 286–287, Jan. 2011.

[110] W. Jeng, D. He, and J. Jiang, "User participation in an academic social networking service: A survey of open group users on Mendeley," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 5, pp. 890–904, May 2015.

[111] H. Alhoori and R. Furuta, "Understanding the dynamic scholarly research needs and behavior as applied to social reference management," in *Proceedings of the 15th international conference on theory and practice of digital libraries*, 2011, vol. 6966, pp. 169–178.

[112] H. Alhoori, C. Thompson, R. Furuta, J. Impagliazzo, E. A. Fox, M. Samaka, and S. Al-Maadeed, "The evolution of scholarly digital library needs in an international environment: Social reference management systems and Qatar," in *Proceedings of the 15th International Conference on Asia-Pacific Digital Libraries*, 2013, vol. 8279, pp. 180–181.

[113] R. Fidel, “Are we there yet?: Mixed methods research in library and information science,” *Libr. Inf. Sci. Res.*, vol. 30, no. 4, pp. 265–272, Dec. 2008.

[114] E. Mohammadi, M. Thelwall, S. Haustein, and V. Larivière, “Who reads research articles? An altmetrics analysis of mendeley user categories,” *J. Assoc. Inf. Sci. Technol.*, 2015.

[115] J. Emanuel, “Users and citation management tools: use and support,” *Ref. Serv. Rev.*, vol. 41, no. 4, pp. 639–659, Nov. 2013.

[116] J. Hallmark, “Access and retrieval of recent journal articles: A comparative study of chemists and geoscientists,” *Issues Sci. Technol. Librariansh.*, vol. 40, 2004.

[117] C. George, A. Bright, T. Hurlbert, E. C. Linke, G. ST Clair, and J. Stein, “Scholarly use of information: Graduate students’ information seeking behaviour,” *Inf. Res.*, vol. 11, no. 4, Jul. 2006.

[118] K. Hoffmann, F. Antwi-Nsiah, V. Feng, and M. Stanley, “Library research skills: A needs assessment for graduate student workshops,” *Issues Sci. Technol. Librariansh.*, vol. 53, 2008.

[119] L. Zach, “When is ‘enough’ enough? Modeling the information-seeking and stopping behavior of senior arts administrators,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 56, no. 1, pp. 23–35, Jan. 2005.

[120] C. W. Choo, “Information seeking in organizations: epistemic contexts and contests,” *Inf. Res.*, vol. 12, no. 2, Jan. 2007.

[121] D. Ellis, “A behavioral approach to information retrieval system design,” *J. Doc.*, vol. 45, no. 3, pp. 171–212, Mar. 1989.

[122] C. M. Brown, “Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists,” *J. Am. Soc. Inf. Sci.*, vol. 50, no. 10, pp. 929–943, 1999.

[123] N. Bakkalbasi, K. Bauer, J. Glover, and L. Wang, “Three options for citation tracking: Google Scholar, Scopus and Web of Science,” *Biomed. Digit. Libr.*, vol. 3, p. 7, 2006.

[124] N. Vibert, J. Rouet, C. Ros, M. Ramond, and B. Deshoullieres, “The use of online electronic information resources in scientific research: The case of neuroscience,” *Libr. Inf. Sci. Res.*, vol. 29, no. 4, pp. 508–532, Dec. 2007.

[125] A. Mas-Bleda, M. Thelwall, K. Kousha, and I. F. Aguillo, “Do highly cited researchers successfully use the social web?,” *Scientometrics*, vol. 101, no. 1, pp. 337–356, Oct. 2014.

[126] P. Jump, “University of Montreal cancels Wiley-Blackwell deal subscription,” Jan-2014. [Online]. Available: https://www.timeshighereducation.co.uk/news/university-of-montreal-cancels-wiley-blackwell-deal-subscription/2010888.article.

[127] P. Zhou, “The growth momentum of China in producing international scientific publications seems to have slowed down,” *Inf. Process. Manag.*, vol. 49, no. 5, pp. 1049–1051, Sep. 2013.

[128] L. Leydesdorff and C. Wagner, “Macro-level indicators of the relations between research funding and research output,” *J. Informetr.*, vol. 3, no. 4, pp. 353–362, Oct. 2009.

[129] L. Bornmann and L. Leydesdorff, “Macro-indicators of citation impacts of six prolific countries: InCites data and the statistical significance of trends,” *PloS ONE*, vol. 8, no. 2, p. e56768, Feb. 2013.

[130] D. A. King, “The scientific impact of nations,” *Nature*, vol. 430, no. 6997, pp. 311–316, Jul. 2004.

[131] R. Smith, “Measuring the social impact of research,” *BMJ*, vol. 323, no. 7312, p. 528, Sep. 2001.

[132] National Science Foundation, “Dissemination and sharing of research results,” 2011. [Online]. Available: http://www.nsf.gov/bfa/dias/policy/dmp.jsp.

[133] “San Francisco declaration on research assessment,” Dec-2012. [Online]. Available: http://www.ascb.org/dora/.

[134] H. Piwowar, “Altmetrics: Value all research products,” *Nature*, vol. 493, no. 7431, p. 159, Jan. 2013.

[135] M. Stebbins, “Expanding public access to the results of federally funded research,” Feb-2013. [Online]. Available: https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.

[136] Higher Education Funding Council for England, “Research Excellence Framework: Decisions on assessing research impact,” Bristol, UK, 2011.

[137] I. Viney, “Altmetrics: Research council responds,” *Nature*, vol. 494, no. 7436, p. 176, Feb. 2013.

[138] T. Bloom, E. Ganley, and M. Winker, “Data access for the open access literature: PLOS’s data policy,” *PLoS Med.*, vol. 11, no. 2, p. e1001607, Feb. 2014.

[139] C. L. Borgman and J. Furner, “Scholarly communication and bibliometrics,” *Annu. Rev. Inf. Sci. Technol.*, vol. 36, no. 1, pp. 2–72, Feb. 2005.

[140] L. Bornmann, “What is societal impact of research and how can it be assessed? A literature survey,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. 2, pp. 217–233, Feb. 2013.

[141] H. F. Moed, *Citation Analysis in Research Evaluation*. Springer Netherlands, 2005.

[142] S. Fausto, F. A. Machado, L. F. J. Bento, A. Iamarino, T. R. Nahas, and D. S. Munger, “Research blogging: Indexing and registering the change in science 2.0,” *PLoS ONE*, vol. 10, no. 4, p. e0124184, Dec. 2012.

[143] Y. Ding, E. K. Jacob, Z. Zhang, S. Foo, E. Yan, N. L. George, and L. Guo, “Perspectives on social tagging,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 12, pp. 2388–2401, Dec. 2009.

[144] E. Adie and W. Roe, “Altmetric: Enriching scholarly content with article-level discussion and metrics,” *Learn. Publ.*, vol. 26, no. 1, pp. 11–17, Jan. 2013.

[145] J. C. Wallis, E. Rolando, and C. L. Borgman, “If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology,” *PLoS ONE*, vol. 8, no. 7, p. e67332, Jul. 2013.

[146] M. Thelwall and N. Maflahi, “Are scholarly articles disproportionately read in their own country? An analysis of mendeley readers,” *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 6, pp. 1124–1135, Jun. 2015.

[147] S. J. Darmoni, F. Roussel, J. Benichou, B. Thirion, and N. Pinhas, “Reading factor: A new bibliometric criterion for managing digital libraries,” *J. Med. Libr. Assoc.*, vol. 90, no. 3, pp. 323–327, Jul. 2002.

[148] P. Vinkler, “Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries,” *Scientometrics*, vol. 74, no. 2, pp. 237–254, Feb. 2008.

[149] F. Moya-Anegón and V. Herrero-Solana, “Science in america latina: A comparison of bibliometric and scientific-technical indicators,” *Scientometrics*, vol. 46, no. 2, pp. 299–320, Oct. 1999.

[150] L. Tasli, N. Kacar, and E. H. Aydemir, “Scientific productivity of OECD countries in dermatology journals within the last 10-year period,” *Int. J. Dermatol.*, vol. 51, no. 6, pp. 665–671, Jun. 2012.

[151] S. A. Meo, A. A. Al Masri, A. M. Usmani, A. N. Memon, and S. Z. Zaidi, “Impact of GDP, spending on R&D, number of universities and scientific journals on research publications among asian countries,” *PLoS ONE*, vol. 8, no. 10, p. e66449, Jun. 2013.

[152] H. Lima, T. H. P. Silva, M. M. Moro, R. L. T. Santos, W. Meira, and A. H. F. Laender, “Aggregating productivity indices for ranking researchers across multiple areas,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital libraries*, 2013, pp. 97–106.

[153] C. Neylon and S. Wu, “Article-level metrics and the evolution of scientific impact,” *PLoS Biol.*, vol. 7, no. 11, p. e1000242, Nov. 2009.

[154] J. Bollen, H. Van de Sompel, A. Hagberg, and R. Chute, “A principal component analysis of 39 scientific impact measures,” *PloS ONE*, vol. 4, no. 6, p. e6022, Jun. 2009.

[155] TPDL 2013, “‘Being cited in a tweet IS a citation’ - Christine Borgman @SciTechProf #Twitter #tpdl2013,” 2013. [Online]. Available: https://twitter.com/tpdl2013/status/382053871444844544.

[156] J. Priem, P. Groth, and D. Taraborelli, “The altmetrics collection,” *PloS ONE*, vol. 7, no. 11, p. e48753, Nov. 2012.

[157] J. Priem, H. A. Piwowar, and B. M. Hemminger, “Altmetrics in the wild: Using social media to explore scholarly impact,” *arXiv:1203.4745*, Mar. 2012.

[158] J. Priem and B. M. Hemminger, “Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web,” *First Monday*, vol. 15, no. 7, Jul. 2010.

[159] J. Priem, D. Taraborelli, P. Groth, and C. Neylon, “Altmetrics: a manifesto,” *October*, 2010. [Online]. Available: http://altmetrics.org/manifesto/.

[160] PLOS, “PLOS article-level metrics (ALMs): measuring the impact of research,” 2012. [Online]. Available: http://article-level-metrics.plos.org/.

[161] Counting Online Usage of Networked Electronic Resources, “Introduction to release 1 of the COUNTER code of practice for usage factors,” 2014. [Online]. Available: http://www.projectcounter.org/usage_factor.html.

[162] G. Baynes, “Article level metrics on nature.com,” Oct-2012. [Online]. Available: http://www.nature.com/press_releases/article-metrics.html.

[163] Springer, “Springer now sharing data from Altmetric on SpringerLink,” Jan-2014. [Online]. Available: http://www.springer.com/about+springer/media/pressreleases?SGWID=0-11002-6-1453458-0.

[164] C. Bower, “Redefining impact – altmetrics now on journals from BMJ,” Oct-2013. [Online]. Available: http://blogs.bmj.com/bmj-journals-development-blog/2013/10/21/redefining-impact-altmetrics-now-on-journals-from-bmj/.

[165] E. Drage, “Altmetrics on CJO – tracking and analysing online activity,” Sep-2013. [Online]. Available: http://blog.journals.cambridge.org/2013/09/altmetrics-on-cjo-tracking-and-analysing-online-activity/.

[166] Altmetric, “Altmetric for Scopus,” 2012. [Online]. Available: http://support.altmetric.com/knowledgebase/articles/83246-altmetric-for-scopus.

[167] “All you can tweet,” *Nat. Chem.*, vol. 5, no. 4, p. 247, Mar. 2013.

[168] M. Thelwall, S. Haustein, V. Larivière, and C. R. Sugimoto, “Do altmetrics work? Twitter and ten other social web services,” *PLoS ONE,* vol. 8, no. 5, p. e64841, May 2013.

[169] S. Haustein, I. Peters, C. R. Sugimoto, M. Thelwall, and V. Larivière, “Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 656–669, Apr. 2014.

[170] X. Shuai, Z. Jiang, X. Liu, and J. Bollen, “A comparative study of academic and Wikipedia ranking,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital libraries*, 2013, pp. 25–28.

[171] L. Waltman and R. Costas, “F1000 recommendations as a potential new data source for research evaluation: A comparison with citations,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 3, pp. 433–445, Mar. 2014.

[172] R. Costas, Z. Zahedi, and P. Wouters, “Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective,” *arXiv:1401.4321*, Jan. 2014.

[173] J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, and J. Terliesner, “Beyond citations : Scholars ’ visibility on the social Web,” in *Proceedings of the 17th International Conference on Science and Technology Indicators*, 2012, pp. 98–109.

[174] E. Mohammadi and M. Thelwall, “Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 8, pp. 1627–1638, Aug. 2014.

[175] Z. Zahedi, R. Costas, and P. Wouters, “How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications,” *Scientometrics*, vol. 101, no. 2, pp. 1491–1513, Nov. 2014.

[176] K. Holmberg and M. Thelwall, “Disciplinary differences in Twitter scholarly communication,” *Scientometrics*, vol. 101, no. 2, pp. 1027–1042, Nov. 2014.

[177] T. Brody, S. Harnad, and L. Carr, “Earlier web usage statistics as predictors of later citation impact,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 57, no. 8, pp. 1060–1072, Jun. 2006.

[178] S. Lawrence, “Online Or invisible?,” *Nature*, vol. 411, no. 6837, p. 521, 2001.

[179] K. Antelman, “Do open-access articles have a greater research impact?,” *Coll. Res. Libr.*, vol. 65, no. 5, pp. 372–382, Sep. 2004.

[180] S. Harnad and T. Brody, “Comparing the impact of open access vs. non OA articles in the same journals,” *D-Lib Mag.*, vol. 10, no. 6, Jun. 2004.

[181] K. Kousha and M. Abdoli, “The citation impact of open access agricultural research: A comparison between OA and non-OA publications,” *Online Inf. Rev.*, vol. 34, no. 5, pp. 772–785, Sep. 2010.

[182] T. Koler-Povh, P. Južnič, and G. Turk, “Impact of open access on citation of scholarly publications in the field of civil engineering,” *Scientometrics*, vol. 98, no. 2, pp. 1033–1045, Feb. 2014.

[183] C. Hajjem, S. Harnad, and Y. Gingras, “Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact,” *IEEE Data Eng. Bull.*, vol. 28, no. 4, pp. 39–47, Jun. 2005.

[184] M. Norris, C. Oppenheim, and F. Rowland, “The citation advantage of open-access articles,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 12, pp. 1963–1972, Oct. 2008.

[185] J. Xia, R. L. Myers, and S. K. Wilhoite, “Multiple open access availability and citation impact,” *J. Inf. Sci.*, vol. 37, no. 1, pp. 19–28, Dec. 2010.

[186] M. J. Mccabe and C. M. Snyder, “Identifying the effect of open access on citations using a panel of science journals,” *Econ. Inq.*, vol. 52, no. 4, pp. 1284–1300, Oct. 2014.

[187] M. J. Kurtz, G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, E. Henneken, and S. S. Murray, “The effect of use and access on citations,” *Inf. Process. Manag.*, vol. 41, no. 6, pp. 1395–1402, Dec. 2005.

[188] G. Eysenbach, “Citation advantage of open access articles,” *PLoS Biol.*, vol. 4, no. 5, p. e157, May 2006.

[189] Y. Gargouri, C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, and S. Harnad, “Self-selected or mandated, open access increases citation impact for higher quality research,” *PLoS ONE*, vol. 5, no. 10, p. e13636, Oct. 2010.

[190] X. Shuai, A. Pepe, and J. Bollen, “How the scientific community reacts to newly submitted preprints: article downloads, Twitter mentions, and citations,” *PLoS ONE*, vol. 7, no. 11, p. e47523, Nov. 2012.

[191] H. G. Allen, T. R. Stanton, F. Di Pietro, and G. L. Moseley, “Social media release increases dissemination of original articles in the clinical pain sciences,” *PLoS ONE*, vol. 8, no. 7, p. e68914, Jul. 2013.

[192] S. Haustein, I. Peters, J. Bar-Ilan, J. Priem, H. Shema, and J. Terliesner, “Coverage and adoption of altmetrics sources in the bibliometric community,” *Scientometrics*, vol. 101, no. 2, pp. 1145–1163, Nov. 2014.

[193] H. Shema, J. Bar-Ilan, and M. Thelwall, “Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics,” *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 5, pp. 1018–1027, May 2014.

[194] P. Jacsó, “Comparison of journal impact rankings in the SCImago journal & country rank and the journal citation reports databases,” *Online Inf. Rev.*, vol. 34, no. 4, pp. 642–657, Aug. 2010.

[195] The World Bank, “World DataBank,” 2014. [Online]. Available: http://databank.worldbank.org/data/home.aspx.

[196] United Nations Statistics Division, “National accounts main aggregates database,” 2014. [Online]. Available: http://unstats.un.org/unsd/snaama/Introduction.asp.

[197] "R&D magazine," 2014. [Online]. Available: http://www.rdmag.com/.

[198] The World Economic Forum, "The global information technology report 2014," 2014.

[199] Google Inc., "Google Scholar Metrics," 2014. [Online]. Available: https://scholar.google.com/citations?view_op=top_venues.

[200] M. Norris, C. Oppenheim, and F. Rowland, "Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR," *Online Inf. Rev.*, vol. 32, no. 6, pp. 709–715, Nov. 2008.

[201] C. Kreibich, "A parser for Google Scholar," 2014. [Online]. Available: http://www.icir.org/christian/scholar.html.

[202] P. M. Davis, "Author-choice open-access publishing in the biological and medical literature: A citation analysis," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 1, pp. 3–8, Jan. 2009.

[203] J. Di Leo, "Against rank," *Inside Higher Ed*, Jun-2010. [Online]. Available: http://www.insidehighered.com/views/2010/06/21/dileo.

[204] P. A. Lawrence, "Rank injustice," *Nature*, vol. 415, no. 6874, pp. 835–836, Feb. 2002.

[205] J. Howard, "New ratings of humanities journals do more than rank they rankle," *The Chronicle of Higher Education*, Oct-2008. [Online]. Available: http://chronicle.com/article/New-Ratings-of-Humanities/29072/.

[206] J. Rowbotham, "End of an ERA: journal rankings dropped," *The Australian*, May-2011. [Online]. Available: http://www.theaustralian.com.au/higher-education/end-of-an-era-journal-rankings-dropped/story-e6frgcjx-1226065864847.

[207] E. Rinia, T. Van Leeuwen, E. Bruins, H. Van Vuren, and A. Van Raan, "Citation delay in interdisciplinary knowledge exchange," *Scientometrics*, vol. 51, no. 1, pp. 293–309, Apr. 2001.

[208] B. González-Pereira, V. P. Guerrero-Bote, and F. Moya-Anegón, "A new approach to the metric of journals' scientific prestige: The SJR indicator," *J. Informetr.*, vol. 4, no. 3, pp. 379–391, Jul. 2010.

[209] C. Bergstrom, "Eigenfactor: Measuring the value and prestige of scholarly journals," *Coll. Res. Libr. News* , vol. 68, no. 5, pp. 314–316, May 2007.

[210] A. Serenko and M. Dohan, “Comparing the expert survey and citation impact journal ranking methods: Example from the field of Artificial Intelligence,” *J. Informetr.*, vol. 5, no. 4, pp. 629–648, Oct. 2011.

[211] C. W. Holsapple, “A publication power approach for identifying premier information systems journals,” *J. Am. Soc. Inf. Sci.*, vol. 59, no. 2, pp. 166–185, Jan. 2008.

[212] Z. Zhuang, E. Elmacioglu, D. Lee, and C. L. Giles, “Measuring conference quality by mining program committee characteristics,” in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital libraries*, 2007, pp. 225–234.

[213] S. Yan and D. Lee, “Toward alternative measures for ranking venues: a case of database research community,” in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital libraries*, 2007, pp. 235–244.

[214] W. S. Martins, M. A. Gonçalves, A. H. F. Laender, and G. L. Pappa, “Learning to assess the quality of scientific conferences,” in *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital libraries*, 2009, pp. 193–202.

[215] E. Rahm and A. Thor, “Citation analysis of database publications,” *ACM SIGMOD Rec.*, vol. 34, no. 4, pp. 48–53, Dec. 2005.

[216] X. Li, M. Thelwall, and D. Giustini, “Validating online reference managers for scholarly impact measurement,” *Scientometrics*, vol. 91, no. 2, pp. 461–471, May 2011.

[217] P. Kraker, C. Körner, K. Jack, and M. Granitzer, “Harnessing user library statistics for research evaluation and knowledge domain visualization,” in *Proceedings of the 21st international conference companion on World Wide Web*, 2012, pp. 1017–1024.

[218] C. Tenopir and D. W. King, *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*. Washington, D.C.: Special Libraries Association, 2000.

[219] L. M. Rudner, M. Miller-Whitehead, and J. S. Gellmann, “Who is reading on-line education journals? Why? And what are they reading?,” *D-Lib Mag.*, vol. 8, no. 12, Dec. 2002.

[220] K. Eason, S. Richardson, and L. Yu, “Patterns of use of electronic journals,” *J. Doc.*, vol. 56, no. 5, pp. 477–504, Oct. 2000.

[221] Mendeley Ltd., “Mendeley API,” 2013. [Online]. Available: http://dev.mendeley.com/.

[222] Wikipedia, “Digital humanities,” 2013. [Online]. Available: https://en.wikipedia.org/wiki/Digital_humanities.

[223] Altmetric LLP, “The Altmetric Explorer,” 2013. [Online]. Available: http://www.altmetric.com/aboutexplorer.php.

[224] S. K.-W. Chu and N. Law, “Development of information search expertise: postgraduates’ knowledge of searching skills,” *portal Libr. Acad.*, vol. 7, no. 3, pp. 295–316, Jul. 2007.

[225] Springer, “Journals subscription, recommend to your librarian / information specialist,” 2015. [Online]. Available: http://www.springer.com/generic/order/journals+subscription?SGWID=0-40514-0-0-0.

[226] G. Perlman, “The HCI bibliography project,” *ACM SIGCHI Bull. - Spec. issue Comput. Support. Coop. Work*, vol. 23, no. 3, pp. 15–20, Jul. 1991.

[227] M. Kuhn and R. Wattenhofer, “The layered world of scientific conferences,” in *Progress in WWW Research and Development*, 2008, pp. 81–92.

[228] R. Gove, C. Dunne, B. Shneiderman, J. Klavans, and B. Dorr, “Evaluating visual and statistical exploration of scientific literature networks,” in *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2011, pp. 217–224.

[229] C. Blake and W. Pratt, “Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 57, no. 13, pp. 1740–1749, Nov. 2006.

[230] D. Shenk, *Data Smog: Surviving the Information Glut*. New York, NY: HarperCollins, 1997.

[231] C. Speier, J. S. Valacich, and I. Vessey, “The influence of task interruption on individual decision making: an information overload perspective,” *Decis. Sci.*, vol. 30, no. 2, pp. 337–360, Mar. 1999.

[232] R. Klamma, P. M. Cuong, and Y. Cao, “You never walk alone : Recommending academic events based on social network analysis,” in *Proceedings of the First International Conference on Complex Science*, 2009, vol. 4, no. 1, pp. 657–670.

[233] H. Luong, T. Huynh, S. Gauch, P. Do, and K. Hoang, “Publication venue recommendation using author network’s publication history,” in *Proceedings on the 4th Asian Conference on Intelligent Information and Database Systems*, 2012, pp. 426–435.

[234] I. Boukhris and R. Ayachi, “A novel personalized academic venue hybrid recommender,” in *IEEE 15th International Symposium on Computational Intelligence and Informatics*, 2014, pp. 465–470.

[235] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, “A clustering approach for collaborative filtering recommendation using social network analysis,” *J. Univers. Comput. Sci.*, vol. 17, no. 4, pp. 583–604, Feb. 2011.

[236] Z. Yang and B. D. Davison, “Venue recommendation: Submitting your paper with style,” in *Proceedings of the 11th International Conference on Machine Learning and Applications*, 2012, vol. 1, pp. 681–686.

[237] E. Medvet, A. Bartoli, and G. Piccinin, “Publication venue recommendation based on paper abstract,” in *Proceedings of the IEEE 26th International Conference on Tools with Artificial Intelligence*, 2014, pp. 1004–1010.

[238] O. Küçüktunç, E. Saule, K. Kaya, and Ü. V. Çatalyürek, “Recommendation on academic networks using direction aware citation analysis,” *arXiv:1205.1143*, May 2012.

[239] O. Kucuktunc, E. Saule, K. Kaya, and U. V Catalyurek, “TheAdvisor: A webservice for ac ademic recommendation,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2013, pp. 433–434.

[240] E. Minkov, B. Charrow, J. Ledlie, S. Teller, and T. Jaakkola, “Collaborative future event recommendation,” in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 819–828.

[241] H. Khrouf and R. Troncy, “Hybrid event recommendation using linked data and user diversity,” in *Proceedings of the 7th ACM conference on Recommender systems*, 2013, pp. 185–192.

[242] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, “Recommending social events from mobile phone location data,” in *Proceedings of the IEEE 10th International Conference on Data Mining*, 2010, pp. 971–976.

[243] P. Resnick and H. R. Varian, “Recommender systems,” *Commun. ACM*, vol. 40, no. 3, pp. 56–58, Mar. 1997.

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

[245] J. Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems,” in *The Adaptive Web*, 2007, pp. 291–324.

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

[247] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, “GroupLens: Applying collaborative filtering to Usenet news,” *Commun. ACM*, vol. 40, no. 3, pp. 77–87, Mar. 1997.

[248] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in *Proceedings of the 14th conference on Uncertainty in Artificial Intelligence*, 1998, pp. 43–52.

[249] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, “Item-based collaborative filtering recommendation algorithms,” in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 285–295.

[250] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, “Recommending and evaluating choices in a virtual community of use,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 194–201.

[251] U. Shardanand and P. Maes, “Social information filtering: Algorithms for automating ‘word of mouth,’” in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1995, pp. 210–217.

[252] A. Woodruff, R. Gossweiler, J. Pitkow, E. H. Chi, and S. K. Card, “Enhancing a digital book with a reading recommender,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2000, pp. 153–160.

[253] M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in *The Adaptive Web*, 2007, vol. 4321, pp. 325–341.

[254] W. Chu and S.-T. Park, “Personalized recommendation on dynamic content using predictive bilinear models,” in *Proceedings of the 18th international conference on World Wide Web*, 2009, pp. 691–700.

[255] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, “Methods and metrics for cold-start recommendations,” in *Proceedings of the 25th annual international*

ACM SIGIR conference on Research and development in information retrieval, 2002, pp. 253–260.

[256] M. Balabanović and Y. Shoham, “Fab: Content-based, collaborative recommendation,” *Commun. ACM*, vol. 40, no. 3, pp. 66–72, Mar. 1997.

[257] S.-T. Park, D. Pennock, O. Madani, N. Good, and D. DeCoste, “Naïve filterbots for robust cold-start recommendations,” in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 699–705.

[258] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, “An algorithmic framework for performing collaborative filtering,” in *Proceedings of the 22nd Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, 1999, pp. 230–237.

[259] L. Bottou and O. Bousquet, “The tradeoffs of large scale learning,” in *Advances in Neural Information Processing Systems*, 2008, vol. 20, pp. 161–168.

[260] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Application of dimensionality reduction in recommender system - a case study,” in *ACM WebKDD Web Mining for E-Commerce Workshop*, 2000.

[261] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[262] D. Parra and P. Brusilovsky, “Collaborative filtering for social tagging systems: an experiment with CiteULike,” in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 237–240.

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

[264] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, “Enhancing digital libraries with TechLens,” in *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital libraries*, 2004, pp. 228–236.

[265] K. Sugiyama and M.-Y. Kan, “Scholarly paper recommendation via user’s recent research interests,” in *Proceedings of the 10th Joint Conference on Digital Libraries*, 2010, pp. 29–38.

[266] N. Agarwal, E. Haque, H. Liu, and L. Parsons, “Research paper recommender systems: a subspace clustering approach,” in *Proceedings of the 6th International Conference on Web-Age Information Management*, 2005, vol. 3739, pp. 475–491.

[267] M. Ohta, T. Hachiki, and A. Takasu, “Related paper recommendation to support online-browsing of research papers,” in *Proceedings of the Fourth International Conference on the Applications of Digital Information and Web Technologies*, 2011, pp. 130–136.

[268] J. Beel, “Link analysis in mind maps : A new approach to determining document relatedness,” in *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication*, 2010, no. 38, pp. 1–5.

[269] Z. Guan, C. Wang, J. Bu, C. Chen, K. Yang, D. Cai, and X. He, “Document recommendation in social tagging services,” in *Proceedings of the 19th international conference on World Wide Web*, 2010, pp. 391–400.

[270] M. Gori and A. Pucci, “Research paper recommender systems: a random-walk based approach,” in *Proceedings of the International Conference on Web Intelligence*, 2006, pp. 778–781.

[271] S. Pohl, F. Radlinski, and T. Joachims, “Recommending related papers based on digital library access records,” in *Proceedings of the 7th ACMIEEECS Joint Conference on Digital libraries*, 2007, pp. 417–418.

[272] B. Gipp, J. Beel, and C. Hentschel, “Scienstein: A research paper recommender system,” in *Proceedings of the International Conference on Emerging Trends in Computing*, 2009, pp. 309–315.

[273] C. Nascimento, A. H. F. Laender, S. Altigran, and M. A. Gonçalves, “A source independent framework for research paper recommendation,” in *Proceeding of the 11th annual international ACMIEEE Joint Conference on Digital libraries*, 2011, pp. 297–306.

[274] K. Uchiyama, H. Nanba, A. Aizawa, and T. Sagara, “OSUSUME: cross-lingual recommender system for research papers,” in *Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation*, 2011, pp. 39–42.

[275] Google Scholar Blog, “Scholar updates: making new connections,” 2012. [Online]. Available: http://googlescholar.blogspot.com/2012/08/scholar-updates-making-new-connections.html.

[276] H. Chen, L. Gou, X. Zhang, and C. L. Giles, “Collabseer : A search engine for collaboration discovery categories and subject descriptors,” in *Proceedings of the 11th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2011, pp. 231–240.

[277] E. Yan and R. Guns, “Predicting and recommending collaborations: An author-, institution-, and country-level analysis,” *J. Informetr.*, vol. 8, no. 2, pp. 295–309, Apr. 2014.

[278] S. Das, P. Mitra, and C. L. Giles, “Ranking authors in digital libraries,” in *Proceedings of the 11th Annual Inernational ACM/IEEE Joint Conference on Digital Libraries*, 2011, pp. 251–254.

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

[280] D. Conry, Y. Koren, and N. Ramakrishnan, “Recommender systems for the conference paper assignment problem,” in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 357–360.

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

[282] T. Strohman, W. B. Croft, and D. Jensen, “Recommending citations for academic papers,” in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 705–706.

[283] Y. Lu, J. He, D. Shan, and H. Yan, “Recommending citations with translation model,” in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 2017–2020.

[284] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, “Context-aware citation recommendation,” in *Proceedings of the 19th international conference on World Wide Web*, 2010, pp. 421–430.

[285] J. Tang and J. Zhang, “A discriminative approach to topic-based citation recommendation,” in *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2009, pp. 572–579.

[286] C. Caragea, A. Silvescu, P. Mitra, and C. L. Giles, “Can’t see the forest for the trees? A citation recommendation system,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2013, pp. 111–114.

[287] Y. Song, L. Zhang, and C. L. Giles, "Automatic tag recommendation algorithms for social recommender systems," *ACM Trans. Web*, vol. 5, no. 1, pp. 1–31, 2011.

[288] S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, T. Kanungo, S. Rajagopalan, A. Tomkins, J. A. Tomlin, J. Y. Zien, H. Road, and S. Jose, "SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation," in *Proceedings of the 12th international conference on World Wide Web*, 2003, pp. 178–186.

[289] Y. Song, Z. Zhuang, H. Li, Q. Zhao, J. Li, W.-C. Lee, and C. L. Giles, "Real-time automatic tag recommendation," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 515–522.

[290] P. A. Chirita, S. Costache, S. Handschuh, and W. Nejdl, "P-TAG : Large scale automatic generation of personalized annotation tags for the web," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 845–854.

[291] S. K. Lam and J. Riedl, "Shilling recommender systems for fun and profit," in *Proceedings of the 13th conference on World Wide Web*, 2004, pp. 393–402.

[292] Apache Software Foundation, "Apache Mahout," 2015. [Online]. Available: http://mahout.apache.org/.

[293] T. Dunning, "Accurate methods for the statistics of surprise and coincidence," *Comput. Linguist.*, vol. 19, no. 1, pp. 61–74, Mar. 1993.

[294] S. Owen, R. Anil, T. Dunning, and and E. Friedman, *Mahout in Action*. Manning Publications, 2011.

[295] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.

[296] S. M. McNee, J. Riedl, and J. A. Konstan, "Being accurate is not enough: How accuracy metrics have hurt recommender systems," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, 2006, pp. 1097–1101.

[297] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.

[298] N. Good, J. Ben Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "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, 1999, pp. 439–446.

[299] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "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*, 1998, pp. 345–354.

[300] G. Shani and A. Gunawardana, "Evaluating recommendation systems," in *Recommender Systems Handbook*, Springer, 2009, pp. 257–297.

[301] D. Hull, S. R. Pettifer, and D. B. Kell, "Defrosting the digital library: Bibliographic tools for the next generation web," *PLOS Comput. Biol.*, vol. 4, no. 10, p. e1000204, Oct. 2008.

[302] D. G. Hendry, J. R. Jenkins, and J. F. McCarthy, "Collaborative bibliography," *Inf. Process. Manag.*, vol. 42, no. 3, pp. 805–825, May 2006.

[303] P. Jacso, "Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster," *Libr. Trends*, vol. 56, no. 4, pp. 784–815, 2008.

[304] P. Heymann, G. Koutrika, and H. G. Molina, "Fighting spam on social web sites: A survey of approaches and future challenges," *IEEE Internet Comput.*, vol. 11, no. 6, pp. 36–45, 2007.

[305] C. Lampe and P. Resnick, "Slash(dot) and burn: Distributed moderation in a large online conversation space," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2004, pp. 543–550.

[306] D. G. Hendry and A. Carlyle, "Hotlist or bibliography? A case of genre on the web," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, 2006, p. 51b.

[307] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2004.

[308] E. Wilde, S. Anand, T. Bucheler, M. Jorg, N. Nabholz, and P. Zimmermann, "Collaboration support for bibliographic data," *Int. J. Web Based Communities*, vol. 4, no. 1, pp. 98–109, Jan. 2008.

[309] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?," in *Proceedings of the International Conference on Web Search and Data Mining*, 2008, pp. 195–206.

[310] E. Santos-Neto, M. Ripeanu, and A. Iamnitchi, “Content reuse and interest sharing in tagging communities,” in *Association for the Advancement of Artificial Intelligence Spring Symposium on Social Information Processing*, 2007.

[311] B. Krause, C. Schmitz, A. Hotho, and G. Stumme, “The anti-social tagger - detecting spam in social bookmarking systems,” in *Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, 2008, pp. 61–68.

[312] T. Bogers and A. Van Den Bosch, “Using language modeling for spam detection in social reference manager websites,” in *Proceedings of the 9th Belgian-Dutch Information Retrieval Workshop*, 2009, pp. 87–94.

[313] J. Sabater and C. Sierra, “Review on computational trust and reputation models,” *Artif. Intell. Rev.*, vol. 24, no. 1, pp. 33–60, 2005.

[314] W. Chen, Q. Zeng, and L. Wenyin, “A user reputation model for a user-interactive question answering system,” *Concurr. Comput. Pract. Exp.*, vol. 19, no. 15, pp. 2091–2103, Oct. 2007.

[315] F. Jin, Z. Niu, Q. Zhang, H. Lang, and K. Qin, “A user reputation model for DLDE learning 2.0 community,” in *Proceedings of the 11th International Conference on Asian Digital Libraries*, 2008, pp. 61–70.

[316] P. J. Windley, D. Daley, B. Cutler, and K. Tew, “Using reputation to augment explicit authorization,” in *Proceedings of the ACM workshop on Digital identity management*, 2007, pp. 72–81.

[317] B. Larsen and P. Ingwersen, “The boomerang effect: Retrieving scientific documents via the network of references and citations,” in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 2002, pp. 397–398.

[318] J. Foster, “Collaborative information seeking and retrieval,” *Annu. Rev. Inf. Sci. Technol.*, vol. 40, no. 1, pp. 329–356, 2006.

[319] M. C. Reddy and B. J. Jansen, “A model for understanding collaborative information behavior in context: A study of two healthcare teams,” *Inf. Process. Manag.*, vol. 44, no. 1, pp. 256–273, Jan. 2008.

[320] A. Karunakaran, M. C. Reddy, and P. R. Spence, “Toward a model of collaborative information behavior in organizations,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. 12, pp. 2437–2451, Dec. 2013.

[321] M. Ge, C. Delgado-Battenfeld, and D. Jannach, “Beyond accuracy: Evaluating recommender systems by coverage and serendipity,” in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 257–260.