UNIVERSITAT DE BARCELONA

Altmetrics in Library and Information Science: coverage of sources and use of social media by authors

Aida Pooladian

Departament de Biblioteconomia, Documentació i Communicació Audiovisual

Programa de doctorat en Informació i Comunicació

Altmetrics in Library and Information Science: coverage of sources and use of social media by authors

A thesis submitted by

Aida Pooladian

In partial fulfilment of the requirements of the University of Barcelona for the degree of Doctor of Philosophy

Supervisor: Dr. Ángel Borrego

Barcelona

May, 2018

Abstract

Altmetrics is a movement that aims to capture new and previously invisible types of impact of scholarly publications on social web platforms such as news sites, Wikipedia, blogs, microblogs, social bookmarking tools and online reference managers. This thesis aims to explore the suitability and reliability of two altmetrics resources: Mendeley, a social reference manager website, and Wikipedia, a free online encyclopedia, written collaboratively by the people who use it. The main objectives of this thesis are to explore the use of reference management software and Wikipedia to estimate the readership of academic literature and to compare the evolution in the coverage of LIS literature by altmetric sources and by citation indexes.

To reach these objectives, three different studies were designed. The general method to obtain the initial dataset was to search the Social Sciences Citation Index to retrieve articles and reviews indexed in the category “Information Science & Library Science”. Each record was then searched in Mendeley to obtain the number of bookmarks of the paper and the academic status of the users. Additonally, we performed a fifteen-month longitudinal study of the evolution of bookmarks in Mendeley. Afterwards, we searched each of these articles in Wikipedia, and retrieved all the entries in which they were cited.

The results of these studies show that the correlation between bookmarks and citations was moderate and the overlap between the most frequently bookmarked and the most frequently cited papers increased over time. A significant share of the bookmarks were made by students and professionals, although the shares of bookmarks made by different categories of users changed as time went by. Reviews were bookmarked more frequently than articles, and papers in English had more bookmarks than papers in any other language.

Also results reveal severe limitations in the use of Wikipedia citations for research evaluation. Lack of standardization and incompleteness of Wikipedia references make it difficult to retrieve them. The number of Wikipedia citations is very low, A significant number of references are cited in biographical entries about the authors of the articles, resulting in a phenomenon of accumulated advantage. Nearly one-third of the Wikipedia citations link to an open access source, although this result is probably an underestimate of open access availability, given the incompleteness of Wikipedia citations.

Resumen

Altmetrics es un movimiento que tiene como objetivo capturar nuevos tipos de impacto, antes invisibles, de las publicaciones académicas en las plataformas de la web social, tales como sitios de noticias, Wikipedia, blogs, microblogs, herramientas de marcadores sociales y gestores de referencias en línea. Esta tesis pretende explorar la pertinencia y fiabilidad de dos recursos altmétricos: Mendeley, un sitio web de gestión de referencia social, y Wikipedia, una enciclopedia online gratuita escrita en colaboración por las personas que la utilizan. Actualmente, Wikipedia es la mayor enciclopedia en línea. Los principales objetivos de esta tesis son investigar el uso del software de gestión de referencia y de Wikipedia para estimar el número de lectores de literatura académica y comparar la evolución en la cobertura de la literatura de Ciencias de la Información y Biblioteconomía (Library and Information Science, por sus siglas en inglés), según fuentes altmétricas y por índices de citas.

Para alcanzar estos objetivos, se diseñaron tres estudios diferentes. El método general para acceder al conjunto de datos consistía en utilizar el *Social Sciences Citation Index* para recuperar revisiones y artículos publicados e indexados en la categoría «Information Science & Library Science». Se buscó cada registro en Mendeley para obtener el número de marcadores del documento y el estatus académico de los usuarios. De manera adicional, realizamos un estudio longitudinal durante quince meses para estudiar la evolución en el número de marcadores. Posteriormente, buscamos cada uno de estos artículos en Wikipedia y recuperamos todas las entradas en las que fueron citados.

Los resultados de estos estudios muestran que la correlación entre los marcadores y las citas era moderada y la coincidencia entre los documentos guardados en marcadores con más frecuencia y los más citados aumentó con el tiempo. Una parte significativa de los marcadores los habían hecho estudiantes y profesionales, aunque el número de marcadores hechos por diferentes categorías de usuarios cambiaron con el paso del tiempo. Las revisiones se guardaban en marcadores con más frecuencia que los artículos, y los documentos en inglés tenían más marcadores que los documentos en cualquier otro idioma.

Los resultados también revelan severas limitaciones en el uso de citas de Wikipedia para la evaluación de investigación. La falta de estandarización y el carácter incompleto de las referencias

de Wikipedia dificultan su recuperación. El número de citas de Wikipedia es muy bajo, y un número significativo de referencias se citan en las entradas biográficas dedicadas a los autores de los artículos, lo que supone una ventaja acumulada. Cerca de un tercio de las citas de Wikipedia se vinculan a una fuente de acceso abierto, aunque probablemente este resultado sea una subestimación de la disponibilidad de acceso a estas fuentes, dado que las citas de Wikipedia están incompletas.

Acknowledgment

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Ángel Borrego for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank Dr. Ernest Abadal, the Dean of the Faculty, for his kindness and also for giving me the opportunity to attend conferences and meet so many interesting people.

I am also grateful to Dr. Mario Pérez-Montoro for his step-by-step accompaniment and guidance during my study and also special thanks to him for proposing Dr. Ángel Borrego as my supervisor, I could not have imagined having a better supervisor and mentor for my PhD study.

I would like to thank my thesis committee members for all of their guidance through this process; your discussion, ideas, and feedback have been absolutely invaluable.

I am deeply grateful to Dr. Andreu Sulé Duesa for his insightful interventions during the early days I arrived to Barcelona, and also try to help me, to in all these years, to become more familiar with Spanish culture and language.

Last but not the least, I would especially like to thank my amazing family for the love, support, and constant encouragement I have gotten over the years. In particular, I would like to thank my parents, Parivash and Ahmad, my brother, Babak, and my sisters, Bita & Neda. You are always there for me and I undoubtedly could not have done this without you.

Thank you very much, everyone!

Table of contents

Overview

Overview

This dissertation endeavours to show the coverage of sources and use of social media by the authors of scholarly outputs in library and information science. The impact of scientific work does not appear just as a citation, it can be also measured as the number of views, downloads, bookmarks or mentions in social media.

In academic and research libraries, methods to measure journals' readership have traditionally included shelving statistics (since journals are usually excluded from loan), document delivery data, participative methods (such as surveys and interviews) and citation counts (Haustein, 2014). The transition from print to electronic journals has increased the amount of data available on the frequency of articles being downloaded. Log files in publishers' servers record journal usage, and initiatives such as the Project COUNTER (projectcounter.org) facilitate the reporting of standardized online usage statistics in a consistent, credible and comparable way across publishers. However, journals' global download usage data are not publicly released by publishers, since this information is commercially sensitive.

Before becoming the standard approach for research evaluation, citation analysis was already used to gather data in studies on information behaviour in academic settings. Analysing the sources cited by scholars in their publications was an unobtrusive way of capturing data on researchers' reading habits that might inform decisions on library collection management. In recent years it has been suggested that altmetrics (i.e. web-based metrics such as views, downloads or mentions in social media) might complement citation analysis in research evaluation procedures.

Some authors have suggested that, in a similar fashion to citation analysis, altmetric indicators might also be used in library settings to explore the information needs and behaviour of library users. The rationale is that, for instance, the number of users who bookmark an article in an online reference management software is a potential indicator of the readership of the article. This approach has already been explored, with promising results (Haustein and Siebenlist, 2011; Borrego and Fry, 2012). The use of altmetric data to estimate journal readership might supplement citation analysis in much the same way as a complement to citation counts in research evaluation. Two of the main limitations of

citation analysis when estimating journal readership also apply to research assessment: that is, citations only capture readership among authors who publish and therefore cite, and they also take a long time to accumulate. Altmetric data can be helpful for estimating readership outside the academic community; another advantage is that they are available shortly after publication, since papers can attract readers before they are cited (Bornmann and Haunschild, 2015; Maflahi and Thelwall, 2016; Thelwall and Sud, 2016).

Today, researchers have moved a great deal of their research activity to the web where academic social networking sites allow them to disseminate, comment and collaborate with colleagues (Van Noorden, 2014). Specifically, reference management software allows scholars to record and share their bibliographic references.

Mendeley is an example of a service of this kind that help scholars manage, tag, cite and share academic papers. Researchers may also use the tool to create a public profile in order to disseminate their publication output and establish links with scholars in their field. Mendeley counts how frequently an article is "read" (i.e. how often it is bookmarked by Mendeley users), thus providing information on how academics interact with scholarly information. Thus, the number of Mendeley users who bookmark a given article or journal potentially indicates its readership size (Li et al., 2011).

Altmetrics are non-traditional metrics proposed as an alternative or a complement to traditional citation impact metrics. Altmetrics cover other aspects of the impact of scientific works, such as the number of views, downloads, bookmarks or mentions in social media. A reference to a scientific article in Wikipedia can also be seen as a metric that partially captures the impact of the article. Contrary to other sources of altmetric data, such as social media, in which the easiness of the process may result in casual sharing of research results, citations in Wikipedia may be indicative of stronger engagement of the user with the article. Among its ''five pillars'', Wikipedia enforces strict editorial guidelines striving ''for verifiable accuracy, citing reliable, authoritative sources'' that ensure quality and standard across all the encyclopedia entries. Citations allow Wikipedia editors to make their contributions verifiable by supporting them with trustworthy external sources, and enable readers to locate further information on topics of interest. Thus, citations in Wikipedia can be considered an indication of the transfer of scholarly output to a wider audience.

This dissertation aims to explore wheather altmetric indicators can be useful to inform on the usage and impact of Library and Information Science (LIS) scholarly literature. The study is underpinned by the following research questions:

- What was Mendeley's coverage in 2015 of the LIS literature published in the previous 20-years (1995-2014)?
- What was the share of the LIS literature bookmarked in Mendeley compared to the share of the same literature cited in Web of Science?
- Did the number of bookmarks in Mendeley vary according to the type of paper (article or review)?
- Did the number of bookmarks in Mendeley vary according to the language of the paper?
- What was the academic status of Mendeley users of LIS literature?
- What were the most widely bookmarked journals in LIS according to Mendeley figures?
- What was the evolution in the coverage of LIS literature in Mendeley compared to the evolution in the number of citations received in Web of Science?
- What was the evolution in the population of users of LIS literature by academic status?
- What was the evolution in the population of users of LIS literature by country of residence?
- Were there any differences in the bookmarking of articles and reviews?
- What was the methodological limitations of counting Wikipedia citations?
- What was is the proportion of LIS literature cited in Wikipedia?
- What were the characteristics of Wikipedia entries that cited LIS literature?
- What was the OA availability of the LIS articles cited in Wikipedia?

The research questions have been addressed through three consecutive studies whose results have been published in three papers. The dissertation itself is structured in seven chapters. It starts with an overview, which presents the main concepts related to the dissertation. It then follows Chapter 1, the literature review, which presents previous

studies in the subject area of the thesis. Then Chapter 2, objectives, presents the main and specific objectives of the dissertation. In Chapters 3 (Paper 1), 4 (Paper 2) and 5 (Paper 3), methods for data collection and data analysis of each study, results and discussion are presented. In Chapter 6, Conclusion, the main findings of each research question in the conducted studies are presented. All three papers are attached at the end of the dissertation.

Chapter 1. Literature Review

Chapter 1. Literature review

1.1 Research evaluation

During the last decades there has been an increasing need to show the impact of research. Scientists hired by governments and industries, professors in universities, PhDs working for foundations or in research centres, etc. build research results that are directly effective on all parts of our life such as the medicines we take, our economic policies, our approaches to marketing, the educational strategies used in our schools, therapeutic strategies for the mentally distressed or the techniques for harnessing energy for industry. However, the results of all investigations are not equally reliable. In other words, many research results are published daily but all of them do not have the same quality and, equally important, there is a budget constraint in access to all this information. The tremendous number of journals being published and the continued increase in the cost of yearly subscriptions have made increasingly difficult for libraries to maintain adequate subscription lists. At the same time, libraries have been facing a marked decrease in budgets, gifts and other forms of financial support (Archambault & Larivière, 2009).

So, by using scientific methods of research evaluation, we have to choose among them. Evaluations of research output and impact are particularly relevant given the emphasis today on accountability and documenting the value of research. Research evaluation is used to provide accountability for public funds and to make decisions on funding allocation. One of the traditional ways that has been used to evaluate research is citation analysis. However, citations can be created for many different reasons (Borgman & Furner, 2002) and because both publishing and citation traditions vary between disciplines, new methods for measuring have emerged. The newest indicators which measure scientific output through social media are collectively called altmetrics.

1.2 Citation analysis

There are lots of motivations for a scientist to write, but one of the most important questions raised at the end of the process is “what is the scientific and social impact of my research publications?” (Grimm & Grimm, 2004). In order to answer this question, there are some metrics that rely on citation counting and journal impact factors as mentioned above, named traditional metrics. Scientometrics measure the inherent value

or impact of information indirectly, via citations. References made by the authors can be assumed to be proof of the use of the information in the publications refereed (Vinkler, 2007). On the other hand, the citedness of publications may be used to represent the extent of use. Seglen (1992) argues that, as long as corrections are made to account for differences across fields, citedness can be a useful indicator of scientific impact. Characterization of scientific journals based on citations and references has a long history. He states that citation analysis is a fair evaluation tool for those scientific sub-fields where the publication in the serial literature is the main vehicle of communication.

Citations are also part of the academic reward system (Merton, 1968), with highly cited authors tending to be recognized as having made a significant contribution to science. Holmberg & Thelwall (2014) have noted that citations are important in scholarly communication: they indicate the use of earlier research in new research, and hence it can be argued that they indicate something about the value of the cited research. But no method is perfect alone and citation analysis is not an exception. As Vinkler (2007) points out, this method has some gaps,

1. Some scientific works are only recognised several years after their publication, while any citation analysis is limited to a predetermined citation window (Lindsey, 1989).
2. Papers that are never cited do not necessarily have zero impact (Seglen, 1997).
3. Negative citations are counted the same way as positive citations (Opthof, 1997).

In addition, citation analysis can be influenced by many factors such as having a large number of friends who cite you or self-citing your own papers excessively. Original research data from an investigation can be reused in new research (Fienberg & Martin, 1985) but some articles with original data are not cited or are cited rarely although their data has been used. This means that some important articles in terms of original data may not be recognized through citation analysis (MacRoberts & MacRoberts, 2010), although the dataset could be considered as a unit for measuring research impact (Sarli & Holmes, 2012). As a result, citation analysis cannot be alone a complete way to evaluate researchers.

1.3 Altmetrics

Galligan & Dyas-Correia (2013) point out that citation accounting and journal impact factors have traditionally been used as a means of ascertaining the value of scholarly work and as a way of filtering out only the most significant and relevant material from the huge volume of academic literature produced. As the volume of material has increased and scholarly communication has moved online, the traditional metrics are failing (Priem et al., 2010). Traditional metrics have generally dealt with journals or articles and have not measured other significant research output like blog posts, slideshows, databases, and other important scholarly outputs. New ways are needed to measure the visibility and impact of research. In this context, social media may generate new ways to measure scientific output (Priem & Hemminger, 2010).

Altmetrics or social media metrics were introduced in 2010 by Prime et al. (2010) as an alternative way of measuring broader research impacts in the social web via different tools. Altmetrics can measure the impact at the journal article level as evidenced through social media activity (Galligan & Dyas-Correia, 2013). As Galloway & Pease (2013) state, altmetrics are the tools that help track a scholar's influence and relevance beyond traditional citation metrics. Altmetrics provide immediate feedback because they rely on real-time data and interactions and can be quantified quickly. Piwowar & Priem (2013) described the benefits of altmetrics in these terms: "Altmetrics provide additional, supplementary information and can balance misleading metrics tied to particular journals. More timely than traditional metrics, altmetrics quickly reveal the impact of recent work and add authority to different types of scholarly products not captured as articles. Altmetrics can capture social media references that escape tradicional metrics and reflect public engagement prompted by scholarly writing."

But both citation counting and altmetric indicators have their own difficulties and deficiencies. As some deficiencies of citation analysis are enumerated above, deficiencies of altmetrics should not be overlooked. Although one of the purposes of altmetrics is measuring research impact beyond academia, it is not easy to determine scholarly and non-scholarly audiences in different platforms (Haustein, 2013). Unlike the traditional indicators, which use the scholarly literature, altmetrics rely on new media that have a more dynamic nature; thus, inconsistency of data is another limitation (Fenner, 2014). Additionally, the durability of data and platforms is another challenge (Liu & Adie, 2013).

The potential for manipulating and gaming altmetrics data is also a serious limitation (Priem, Parra, Piwowar, Groth, & Waagmeester, 2012) which is rooted in the lack of quality control on the social web. The majority of new metrics are more appropriate for recent publications and less suitable for old papers. Additionally, altmetrics are prone to biases towards scholars with more Web visibility, who are mainly younger (Priem, 2014). For instance, people who have more friends in the social networks or those who are more active tweeting have a greater chance of being seen or getting more tweets. Moreover, Kwak, Lee, Park & Moon (2010) showed that once retweeted, a tweet gets retweeted almost instantly on the 2nd, 3rd, and 4th hops away from the source, signifying fast diffusion of information after the first retweet.

Finally, the behaviours of scholars in social media are not similar across disciplines, countries and languages, and therefore the normalization of altmetrics for different contexts needs to be considered (Wouters & Costas, 2012).

1.3.1Coverage of altmetric sources

1.3.1.1 Mendeley

As we know, coverage determines the value of databases. Web of Science and Scopus are the databases that have been traditionally used for citation analysis studies. Web of Science, as an important reference for citation analysis, excludes most research contributions published in non-listed journals, books, book chapters, conference proceedings, and most new internet-based outlets. Furthermore, journal articles published in languages other than English are also significantly underrepresented in the ISI-listed journals. In addition, measuring and evaluating research outputs raise particular challenges when one looks at professional schools like, for example, business schools which try to achieve a balance between fostering research excellence and engagement in professional practice. Assessments based on the Web of Science database might underestimate the performance of Business school scholars because it does not take into account most of their contributions disseminated in outlets not included in the Web of Science database. Similarly, assessments using Web of Science might also underestimate the performance of scholars working in non-English language schools (Amara & Landry, 2012).

Scopus indexes a larger number of journals than Web of Science, and includes more internationally diverse and open access journals (Bakkalbasi, Bauer, Glover, & Wang, 2006; Mongeon & Paul-Hus, 2016). The the coverage of active scholarly journals in WoS is 13,605 journals and Scopus is 20,346 journals (Mongeon & Paul-Hus, 2016). Of the 28 million records in Scopus over than 90% is a description of an article in a journal. The number of non-journal sources (books, reports, book series, conference, papers, etc.), at just under thirty thousand, is comparatively small (Sieverts, 2006).

If we are thinking about using altmetric indicators, we need to know the coverage of the sources of altmetric data. Several studies have focused on this question. Based on previous research conducted, Mendeley is the social network with the highest coverage among altmetric sources (62.6%) for 20,000 random publications indexed in the Web of Science (Zahedi, Costas, & Wouters, 2014). Furthermore, Mohammadi & Thelwall found that 44% of social science articles and 13% of the humanities papers from the Web of Science in the year 2008 were covered by Mendeley (Mohammadi, Thelwall, Haustein, & Larivière, 2015). In another study, Li, Thelwall, & Giustini (2011) searched Mendeley, CiteULike and Web of Science for 1,613 Nature and Science articles published in 2007. They found that 92% of the sampled articles had been bookmarked by at least one Mendeley user, and 60% by one or more CiteULike users. Bar-Ilan et al. (2012) studied the publication lists of 57 scientometricians and found that Mendeley covered 82 % of the 1,136 publications of these scientometricians found on Scopus and the correlation between Scopus citations and Mendeley readership counts was 0.448. Bar-Ilan (2012) showed that more than 97 % of the articles published in *Journal of the Association for Information Science and Technology* in the years 2001 to 2011 were bookmarked by Mendeley readers (Thelwall, 2013).

Recently, Zahedi, Costas and Wouters (2017) showed that 86.5% of all the publications are covered by Mendeley and have at least one reader. Also, the share of publications with Mendeley Readership Score is increasing from 84% in 2004 to 89% in 2009, and decreasing from 88% in 2010 to 82% in 2013. However, it is noted that publications from 2010 onwards exhibit on average a higher density of readership versus citation scores. This indicates that compared to citation scores, readership scores are more prevalent for recent publications and hence they could work as an early indicator of research impact.

These findings highlight the potential and value of Mendeley as a tool for scientometric purposes and particularly as a relevant tool to identify highly cited publications. Alos, Thelwall (2017) mentioned that Mendeley reader counts tend to correlate strongly and positively with citation counts within scientific fields, an understanding of causes of citation-reader anomalies is needed before Mendeley reader counts can be used with confidence as indicators.

In response, in this study proposes a list reasons for anomalies based upon an analysis of articles that are highly cited but have few Mendeley readers, or vice versa. The results show that there are both technical and legitimate reasons for differences, with the latter including communities that use research but do not cite it in Scopus-indexed publications or do not use Mendeley. The results also suggest that the lower of the two values (citation counts, reader counts) tends to underestimate of the impact of an article and so taking the maximum is a reasonable strategy for a combined impact indicator.

Thelwall (2017) recently in another article that assesses the total number and proportions of student readers of academic articles in Mendeley across 12 different subjects. The results suggest that whilst few students read mathematics research articles, in other areas, the number of student readers is broadly proportional to the number of research readers. Although the differences in the average numbers of undergraduate readers of articles varies by up to 50 times between subjects, this could be explained by the differing levels of uptake of Mendeley rather than the differing educational value of disciplinary research. Overall, then, the results do not support the claim that journal articles in some areas have substantially more educational value than average for academia, compared with their research value (Thelwall, 2017).

Furthermore, in another study, researchers showed that the overall percentage for unique Google Scholar Citations outside the Web of Science were 73% and 60% for the articles published in *Journal of the Association for Information Science and Technology* and *Scientometrics*, respectively. An important conclusion is that in some subject areas where wider types of intellectual impact indicators outside the Web of Science and Scopus databases are needed for research evaluation, altmetrics can be used to help monitor research performance (Kousha, Thelwall, & Rezaie, 2010).

In addition to the coverage of altmetric sources, in terms of percentage of articles included in these platforms, we need to know who are the creators of this data, i.e. who are the users of platforms used to compile altmetric data. According to the results of previous research, the majority of readers for Mendeley are PhD students, postgraduates and postdocs but other types of academics are also represented (Mohammadi et al., 2014). Nández & Borrego (2013) have pointed to out that half of the respondents to a survey of Academia.edu users were academics (50%), including full professors, permanent, non-permanent and part-time lecturers. However, it should be considered that professions are self-reported and it is possible that, for example, some of the people recorded as Professor might not be full professors (Mohammadi et al., in press). Nández & Borrego (2013) also noted that the respondents were from all age ranges, with highest presence of those in the 31-40 age group (37%) whereas just 14% were older than 51.

1.3.1.2 Wikipedia

As explained above, altmetrics are non-traditional metrics proposed as an alternative or a complement to traditional citation impact metrics. Altmetrics cover other aspects of the impact of scientific works, such as the number of views, downloads, bookmarks or mentions in social media. A reference to a scientific article in Wikipedia can be seen as a metric that partially captures the impact of the article.

Contrary to other sources of altmetric data, such as social media, in which the easiness of the process may result in casual sharing of research results, citations in Wikipedia may be indicative of stronger engagement of the user with the article. Among its ‘‘five pillars’’, Wikipedia enforces strict editorial guidelines striving ‘‘for verifiable accuracy, citing reliable, authoritative sources’’ that ensure quality and standard across all the encyclopedia entries. Citations allow Wikipedia editors to make their contributions verifiable by supporting them with trustworthy external sources, and enable readers to locate further information on topics of interest. Thus, citations in Wikipedia can be considered an indication of the transfer of scholarly output to a wider audience.

Nielsen (2007) was one of the first authors to examine citations in Wikipedia to articles in scholarly journals. He observed that Wikipedia citations correlated strongly with the total number of citations to a journal, but more weakly with the journal’s impact factor. Wikipedia contributors also showed a slight tendency to cite articles in high-impact

journals such as *Nature and Science*. A similar trend was described by Stankus and Spiegel (2010), who observed that both titles topped the list of Wikipedia journal sources for entries on the brain and behavioural sciences. However, the results are different in disciplines with distinct citing behaviours. Thus, Luyt and Tan (2010) found that most citations in a set of Wikipedia history entries were to books, with very few citations of academic journal material. Similarly, Halfaker and Taraborelli (2015) analysed the presence of ISBN, PubMed, DOI and arXiv identifiers in Wikipedia and found that most matches were to books and monographs. To sum up, citations in Wikipedia of scholarly literature have been used as proxy measurements of the encyclopedia's reliability, and differences in verifiability across topics have been identified (Mesgari et al. 2015).

Using a different approach, Huvila (2010) conducted a survey on Wikipedia editors' information behaviour, identifying five groups of contributors who use different information sources. The results indicated a preference among contributors for sources that are available online, although a significant proportion of the original information was based on printed literature, personal expertise and other non-digital sources of information.

Finally, another line of inquiry has explored Wikipedia as an alternative source of evidence about the impact of research. Thus, Evans and Krauthammer (2011) searched PubMed IDs and DOIs in Wikipedia and observed that these articles have higher citation counts than an equivalent random article subset. The fact that articles were cited in Wikipedia soon after publication suggested that Wikipedia citations might represent a resource for assessing articles' impact. This opinion was shared by one-third of the bibliometricians who attended the 17th International Conference on Science and Technology Indicators (STI2012), who believed that the number of Wikipedia links or mentions of an article could be of use in author or article evaluation (Haustein et al. 2014).

Interest in Wikipedia as a source of altmetric data has grown in recent years. In February 2015, Altmetric.com, a start-up focused on tracking and analysing online activity relating to scholarly literature, announced that any mentions of articles and academic output in Wikipedia would be reflected in a new Wikipedia tab on the Altmetric details. In order to capture this information, the academic output that was mentioned had to be referenced with proper Wikipedia citation tags. However, exploratory research led to doubts about

the use of Wikipedia as a source of evidence of the impact of research. Lin and Fenner (2014) found that just 4% of PLOS articles had been cited in Wikipedia.

Thelwall (2016) analysed the presence of astronomy and astrophysics research in Wikipedia, and indicated that the use of Wikipedia citations as a proxy for public interest in research articles was limited, due to the intermediate role of Wikipedia contributors.

Consequently, references reflect the interest of a small number of researchers and amateurs who are enthusiastic Wikipedia editors, rather than the general public. Subsequently, Kousha and Thelwall (2017) showed that only 5% of the articles indexed by Scopus between 2005 and 2012 had been cited in Wikipedia, although this percentage rose to 8% when reviews were considered. In contrast, 33% of the academic monographs indexed by Scopus had attracted at least one Wikipedia citation. They concluded that Wikipedia citations were not common enough to be used for impact assessment of articles in most fields. They also tried to investigate the extent to which academics exploit journalism using content and citation analyses of online *BBC News* stories cited by Scopus articles. A total of 27,234 Scopus-indexed publications have cited at least one *BBC News* story, with a steady annual increase. Citations from the arts and humanities (2.8% of publications in 2015) and social sciences (1.5%) were more likely than citations from medicine (0.1%) and science (<0.1%). Surprisingly, half of the sampled Scopus cited science and technology (53%) and medicine and health (47%) stories were based on academic research, rather than otherwise unpublished information, suggesting that researchers have chosen a lower-quality secondary source for their citations. Nevertheless, the *BBC News* stories that were most frequently cited by Scopus, Google Books, and Wikipedia introduced new information from many different topics, including politics, business, economics, statistics, and reports about events. Thus, news stories are mediating real-world knowledge into the academic domain, a potential cause for concern (Kousha and Thelwall, 2017). More recently, Teplitskiy et al. (2017) analysed whether journals' impact factor and open access (OA) availability were related to their presence in Wikipedia. They found that a journal's impact factor predicts its appearance in Wikipedia, and that its accessibility increases the odds of being referenced in Wikipedia, although to a lesser extent.

1.3.2 Data collection, analysis and interpretation of altmetric indicators

It should be noted that altmetric data are often readily available for any researcher to download and use. In contrast to citation databases such as Web of Science or Scopus where an expensive license is needed in order to access the material, many providers permit data to be reused and integrated in services like Altmetrics.com or Impact Story. The availability of data makes possible for larger groups of researchers to access metrics on their own "impact" as well as that of others. However, as pointed out by Wouters & Costas (2012), many of the services used for altmetric analyses are only partly open, as we know very little about the inner working of commercial companies such as Twitter or Mendeley (Bosman et al., 2006).

Although the impact factor was conceived as a way to assess journals, it is now often used to establish the value of the articles published in those journals, and by extension the quality of individual scientists' work (Fuyuno & Cyranoski, 2006), so it plays a major role in the evaluation of scientific works. However, we have to pay attention to the fact that it is aimed to build journal rankings and measures a journal's average citations. Furthermore, using journal impact factors to measure the quality of individual articles is flawed if citations are not uniformly spread between articles (Weale, Bailey & Lear, 2004). On the other hand, altmetrics can take the form of article downloads or saves, tweets, or simply article views. But if an article has been viewed or downloaded several times, it does not necessarily mean that the article was effective.

Despite the existence of a large body of literature on the meaning of citations, it is difficult to agree on the interpretation of the results of studies based on citation analysis. However, it is even more difficult to interpret altmetric indicators such as the number of tweets or mentions in blogs received by an article. Most of the work done so far has been trying to find out if there is a correlation between traditional metrics, including the number of citations received by an article, and new altmetrics.

The highest correlations between citations and Mendeley readership counts were found for types of users that often authored academic papers, except for associate professors in some sub- disciplines. This suggests that Mendeley readership can reflect usage similar

to traditional citation impact, if the data is restricted to readers who are also authors, without the delay of impact measured by citation counts. At the same time, Mendeley statistics can also reveal the hidden impact of some research papers, such as educational value for non- author users inside academia or the impact of research papers on practice for readers outside academia (Mohammadi et al., 2014). Li, Thelwall, & Giustini (2011) gathered data on how scholarly papers are cited on the web through different search engines. They found statistically significant correlations between traditional and web-based citation counts and reinforce the idea that the web is a rich source of information for measuring scholarly impact. They also investigated how bookmarks in Mendeley and CiteULike reflect papers' scholarly impact: Mendeley bookmarks showed moderate correlation to Web of Science (r=.55) and Google Scholar (r=.60) citations, with CiteULike correlations somewhat lower (Web of Science: .34, Google Scholar: .39), perhaps due to sparser data. The authors concluded that social bookmarking systems are valuable sources for measuring research impact from the readers' point of view.

However, web mentions can be trivial, such as contents lists of journals rather than reference lists of publications. In comparison, Google Scholar works more like an extended bibliometric database that contains many types of publications, like preprints, conference papers, theses and books (Google, 2010), that are outside Web of Science and Scopus and can therefore reflect a wider type of scholarly impact. Brody, Harnad & Carr (2006) found a significant and sizeable correlation between the citation and download impact of articles in physics (0.462), as well as in other arXiv fields: mathematics (0.347), astrophysics (0.477) and condensed matter (0.330). They believe that citation and download counts are just the first two terms in what will be a rich and diverse multiple regression equation predicting and tracking research impact. By comparing the citation counts from Web of Science and Scopus with five online sources of citation data including Google Scholar, Google Books, Google Blogs, PowerPoint presentations and course reading lists, Kousha, Thelwall & Rezaie (2010) proved that online citations are sufficiently numerous to be useful for the impact assessment of research and they also found significant correlations between conventional and online impact indicators, confirming that both assess something similar in scholarly communication. Meho & Yang (2007) showed that Scopus significantly alters the relative ranking of scholars that appear in the middle of the rankings and that Google Scholar stands out in its coverage of conference proceedings as well as international, non-English language journals. The use

of Scopus and Google Scholar, in addition to Web of Science, helps reveal a more accurate and comprehensive picture of the scholarly impact of authors. Prime et al. (2010) proved that correlation and factor analysis suggest that citation and altmetrics indicators track related but distinct impacts, with neither is able to describe the complete picture of scholarly use alone. There are moderate correlations between Mendeley and Web of Science citation, but many altmetric indicators seem to measure impact mostly orthogonal to citation. In summary, the majority of researchers agree with the existence of a correlation between citations and social networks readership counts, however it is low.

1.4 Conclusions

The question remains as to whether altmetrics can be useful to help solve the problem of research evaluation. Nobody can answer this question clearly yet. Based on the results discussed above, as a small sample of the difficulties in both citation analysis and altmetrics, we can say that none of these indicators are complete and it can be more helpful to combine both. Altmetrics complement, and correlate significantly with, traditional measures. Therefore, in future evaluations of the social impact of articles, we should consider not only traditional metrics but also active altmetrics (Liu et al., 2013). Metrics based on these activities could inform broader, faster measures of impact, complementing traditional citation metrics. Of course, we must consider the fact that altmetrics is one of the new concept and it is still possible to work and progress on this subject area, but we should not deny the fact that social network citations are much faster than traditional citations, with 40% occurring within one week of the cited resource's publication.

Chapter 2. Objectives

Chapter 2. Objectives

2.1. Main objectives

The dissertation has three main objectives to be reached in three different studies:

- First, to explore the use of reference management software to estimate the readership of academic literature, taking LIS as a case study. (Study 1).
- Second, to compare the evolution in the coverage of LIS literature, altmetrics sources and by citation indexes. (Study 2).
- Third, to explore the coverage of Library and Information Science (LIS) literature in Wikipedia (Study 3).

2.2. Specific objectives

To reach the three goals stated above, each study has its specific objectives which are listed below.

Regarding to the first main objective, which is to explore the use of reference management software to estimate the readership of academic literature, this dissertation more specifically examines:

- The Mendeley's coverage in 2015 of the LIS literature published in the previous 20-year period (1995-2014).
- The share of the LIS literature bookmarked in Mendeley compared to the share of the same literature cited in Web of Science.
- The number of bookmarks in Mendeley according to the type of paper (article or review)
- The number of bookmarks in Mendeley according to the language of the Paper.
- The academic status of Mendeley users of LIS literature.
- The most widely bookmarked journals in LIS according to Mendeley.

Regarding to the second main objective, which is to compare the evolution in the coverage of LIS literature, altmetrics sources and by citation indexes, this dissertation more specifically examines:

- The evolution in the coverage of LIS literature in Mendeley compared to the evolution in the number of citations received in WoS.
- The evolution in the population of users of LIS literature by academic status.
- The evolution in the population of users of LIS literature by country of residence.
- The differences in the bookmarking of articles and reviews.

Regarding to the third main objective, which is to explore the coverage of Library and Information Science (LIS) literature in Wikipedia, this dissertation more specifically examines:

- The methodological limitations of counting Wikipedia citations.
- The proportion of LIS literature cited in Wikipedia.
- The characteristics of Wikipedia entries that cite LIS literature.
- The OA availability of the LIS articles cited.

Chapter 3. Twenty years of readership of library and information science literature under Mendeley's microscope (Study 1)

Chapter 3. Twenty years of readership of library and information science literature under Mendeley's microscope

Altmetric indicators have been proposed as a complement to citation counts in research evaluation. Conceivably, they might also be used to supplement other methods for estimating journal readership. As mentioned in Chapter 2, the purpose of this study is to explore the use of Mendeley reference management software to characterize the features of the readership of library and information science (LIS) literature. In this Chapter, you can find the methods used for data collection and data analysis, sources used, results, conclusion and discussion of Study 1.

3.1. Methods and data collection

In the first quarter of 2015 we used the Social Sciences Citation Index to retrieve a total of 54,778 papers published between 1995 and 2014 and indexed in the category "Information Science & Library Sciences". The search was limited to articles and reviews, and our analysis excluded all other types of document, such as book reviews, editorials, letters or proceedings. Web of Science defines articles as "reports of research on original works" whereas a review is "a renewed study of material previously studied" (Thomson Reuters, 2013). Throughout this article we use the term "papers" to refer to the set of both articles and reviews.

When this initial search was completed, between 27 April and 8 May 2015 each retrieved record was searched in Mendeley using Webometric Analyst software (Thelwall, 2009) with a query containing the title of the paper, the first author's last name, the year of publication and the digital object identifier.

The software was unable to search 123 records, probably because they were incomplete or contained errors generated while being downloaded from Web of Science.
These records were subsequently removed from the analysis, leaving a final sample of 54,655 papers. For each paper, the journal, language and year of publication were obtained from Web of Science and the number of users and their academic status were recorded from Mendeley. (Note that Mendeley only returns the three main academic status categories for the users of each paper).

3.2. Data analysis and Results

3.2.1. Mendeley coverage of LIS literature

Of the 54,655 LIS papers published between 1995 and 2014, 33,295 (61%) had been bookmarked in Mendeley by at least one user by May 2015 (Table 1). Since the first public beta version of Mendeley was released in August 2008, it was expected that the coverage of the literature would be related to its publication year. Thus, as shown in Figure 1, nearly two-thirds (64%) of the "older" papers — those published between 1995 and 1999 — had no bookmarks. However, the number of articles and reviews with and without bookmarks began to balance in documents published between 2000 and 2004, and the number of papers bookmarked then continued to increase, so that more than 70% of the papers published in the last decade (2005–2014) were bookmarked at least once in Mendeley by May 2015.

Table 1. Mendeley coverage of LIS literature published between 1995 and 2014

	Papers	**% of papers**
Bookmarked	33,295	60.9
Without bookmarks	21,360	39.1
Total	**54,655**	**100.0**

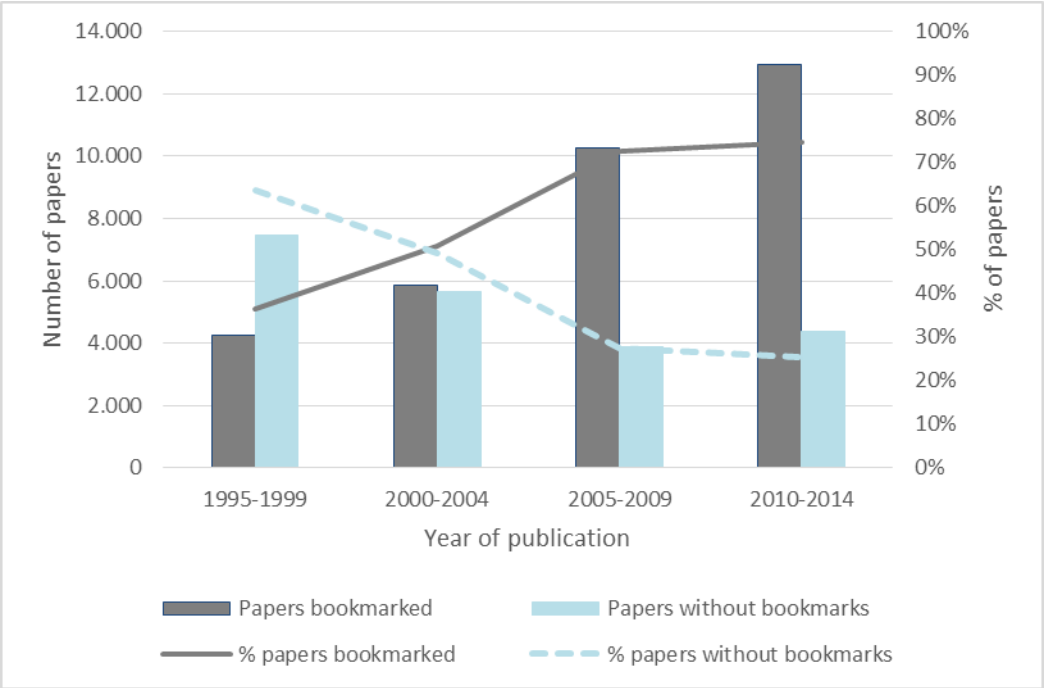

Figure 1. Mendeley coverage of LIS literature by year of publication

One of the possible advantages of bookmark counts for estimating readership is that they may be able to capture usage immediately after publication, whereas citations need much more time to accrue. Figure 2 compares the share of LIS literature bookmarked in Mendeley with the share of papers cited in Web of Science by year of publication. Whereas the percentage of cited papers is higher among papers published between 1995 and 2004, the share of cited and bookmarked papers becomes progressively balanced for papers published between 2005 and 2009. In the case of papers published in the last five years (2010-2014), 55% had been cited in Web of Science at the time of data collection, in May 2015, whereas 75% had been bookmarked at least once in Mendeley.

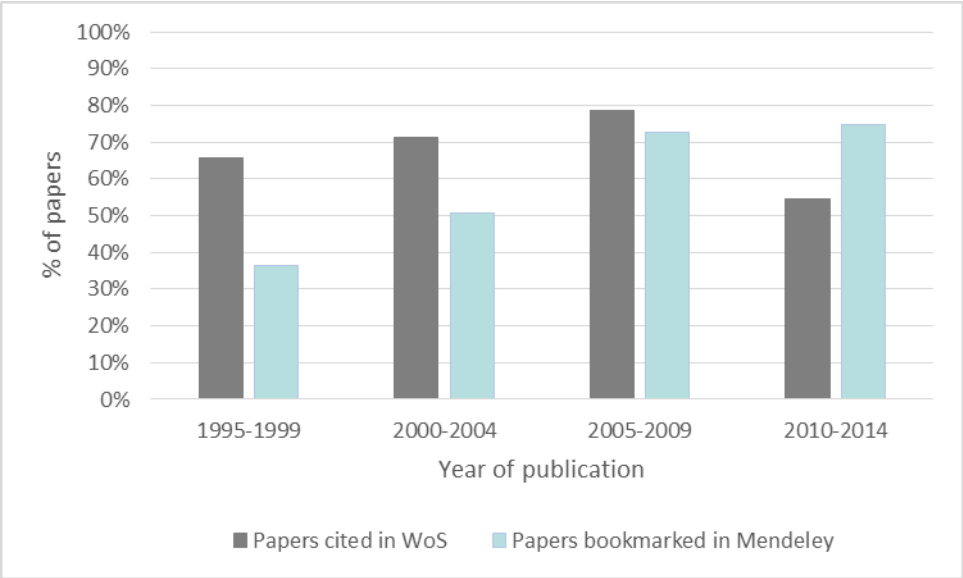

Figure 2. LIS literature cited and bookmarked by year of publication

For each paper, Mendeley provides the number of "readers", i.e., the number of users who bookmarked the paper in their libraries. As shown in Table 2, one quarter of the papers (26%) had between one and five users and over half (56%) had between one and fifteen. At the other end of the scale, an article entitled "Social network sites: definition, history, and scholarship" published in 2007 had 10,217 users. However, this case was particularly extreme, since the second article in the ranking had 893 users.

Table 2. Number of users per paper

	Papers	**% of papers**
1–5 users	8,537	25.6
6–10 users	5,742	17.2
11–15 users	4,380	13.2
16–20 users	3,198	9.6
21–25 users	2,389	7.2
26–30 users	1,701	5.1
31–35 users	1,338	4.0
36–40 users	1,021	3.1
41–45 users	859	2.6
46–50 users	688	2.1
> 50 users	3,442	10.3
Total	**33,295**	**100.0**

The distributions of both bookmarks and citations among papers were skewed, since the top 20% of the papers accounted for 75% of the bookmarks and 82% of the citations (Figure 3). However, the top papers by number of bookmarks and by number of citations are not necessarily the same. If we compare the core articles by number of bookmarks (i.e. those with 21 or more bookmarks each) and the core articles by number of citations (i.e. those with 10 or more citations each), we observe an overlap of 53%. This means that nearly half of the articles among those with the highest number of bookmarks were not in the top by number citations and vice versa.

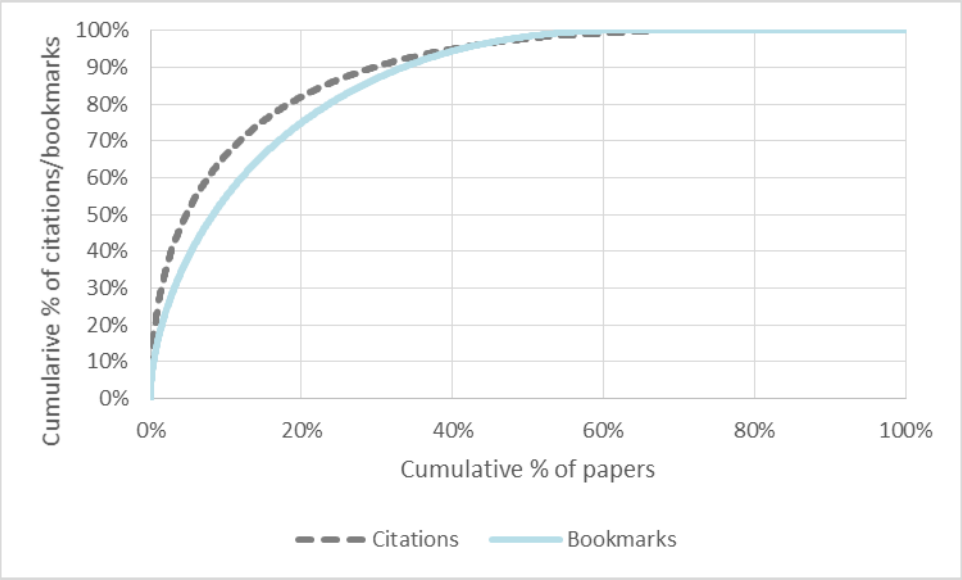

Figure 3. Distributions of citations and bookmarks

3.2.2. Readership by type of paper (article or review)

In total, 67 per cent of the LIS reviews published between 1995 and 2014 were bookmarked in Mendeley, against 61 per cent of the articles. The median and the average number of bookmarks per review were 2.5 times higher than the number of bookmarks per article (Table III). These differences were similar to those observed regarding citations, with a larger share of reviews being cited (81 per cent) compared to articles (66 per cent). In this case the median and the average number of citations per review were four times the median and the average number of citations per article.

Table 3. Bookmarks and citations by type of paper (article or review)

	Articles (n=53,276)	**Reviews (n=1,379)**
Bookmarks		
Papers bookmarked	32,370 (60.8%)	925 (67.1%)
Q1 (bookmarks)	0	0
Median (bookmarks)	4	10
Q3 (bookmarks)	17	39
Average (bookmarks)	13.6	34.0
Standard deviation (bookmarks)	34.0	67.0
Citations		
Papers cited	35,349 (66.4%)	1,116 (80.9%)
Q1 (citations)	0	1
Median (citations)	2	8
Q3 (citations)	7	27
Average (citations)	8.2	33.6
Standard deviation (citations)	28.6	98.6

3.2.3. Readership by language

The number of bookmarks varied according to the language of the paper. As shown in Table 4, the overwhelming majority (95%) of the LIS papers indexed in the Social Sciences Citation Index between 1995 and 2014 were in English. The second language by number of papers was Spanish. Overall, the percentages of English and Spanish papers bookmarked in Mendeley were similar. The number of LIS papers in other languages indexed in the Social Sciences Citation Index was much smaller and, in all cases, the percentage of papers bookmarked in Mendeley and the average and median number of bookmarks per paper were smaller than among English and Spanish papers.

Table 4. Readership by language

Language	Papers	Papers bookmarked	% of papers bookmarked	Average number of bookmarks (SD)	Median number of bookmarks
English	51,912	32,360	62.3	14.7 (53.7)	4
Spanish	1,041	720	69.2	8.6 (15.7)	4
German	716	42	5.9	0.1 (0.4)	0
Portuguese	548	68	12.4	0.6 (3.5)	0
Hungarian	162	21	13.0	0.3 (1.3)	0
French	150	54	36.0	1.2 (2.4)	0
Japanese	118	27	22.9	0.7 (1.7)	0
Other	8	3	37.5	4.1 (9.7)	0
Total	**54,655**	**33,295**	**60.9**	**14.2 (52.5)**	**4**

3.2.4. Academic status of the users of LIS literature

The largest group of users of LIS literature in Mendeley were PhD students (34 per cent), followed by postgraduate students (22 per cent) and librarians (8 per cent), as shown in Table V.

Table 5. Users by academic status

Occupation provided by the Mendeley API	Merged	%
PhD Student	PhD Student	33.8
Doctoral Student		
Student (Postgraduate)	Postgraduate student	22.2
Student (Master)		
Librarian	Librarian	7.5
Assistant Professor	Assistant Professor	6.9
Lecturer		
Researcher (at an Academic Institution)	Researcher (at an Academic Institution)	5.7
Associate Professor	Associate Professor	5.3
Senior Lecturer		
Student (Bachelor)	Student (Bachelor)	5.0
Professor	Professor	4.2
Postdoc	Postdoc	4.2
Other Professional	Other Professional	3.3
Researcher (at a non-Academic Institution)	Researcher (at a non-Academic Institution)	2.0

3.2.5. Readership by journal

The papers in the sample had been published in 124 journals. Nearly all of them had articles bookmarked, although there were notable differences in the percentage of papers with and without bookmarks (Table VI). In the case of one-fifth (19 per cent) of the journals, nearly all the papers had been bookmarked at least once. However, in the case of one-quarter of other journals, a much smaller share of the papers (less than 20 per cent) had been bookmarked.

Table 7 details the top 23 journals by percentage of papers bookmarked. The full list is in Annex 1.

Table 6. Readership by journal

% of papers with users	Number of journals	% of journals
91–100%	23	18.5
81–90%	14	11.3
71–80%	14	11.3
61–70%	6	4.8
51–60%	7	5.6
41–50%	8	6.5
31–40%	9	7.3
21–30%	12	9.7
11–20%	15	12.1
1–10%	14	11.3
0%	2	1.6
Total	**124**	**100**

Table 7. Top journals by percentage of papers bookmarked

Journal	Papers published	Papers bookmarked	% of papers bookmarked	JCR quartile 2014
Information Technology for Development	111	111	100.0	Q3
Telematics and Informatics	152	151	99.3	Q2
International Journal of Computer-Supported Collaborative Learning	167	165	98.8	Q1
Information Systems Journal	362	354	97.8	Q1
Information & Management	1,097	1,065	97.1	Q1
Library & Information Science Research	464	450	97.0	Q2
Information Systems Research	595	576	96.8	Q1
Information and Organization	88	85	96.6	Q1
Journal of the American Society for Information Science and Technology	2,041	1,958	95.9	Q1
International Journal of Information Management	865	829	95.8	Q1
Information Processing & Management	1,288	1,230	95.5	Q2
Journal of Strategic Information Systems	290	276	95.2	Q1
Knowledge Management Research & Practice	225	213	94.7	Q3
Ethics and Information Technology	151	142	94.0	Q2
International Journal of Geographical Information Science	1,179	1,107	93.9	Q1
Journal of Informetrics	451	422	93.6	Q1
Scientometrics	2,826	2,632	93.1	Q1
Government Information Quarterly	710	661	93.1	Q1
Journal of Knowledge Management	361	335	92.8	Q1
Telecommunications Policy	1,098	1,006	91.6	Q2
Information Development	176	161	91.5	Q3
Serials Review	242	221	91.3	Q3
Journal of Documentation	654	596	91.1	Q2

3.3. Discussion

Studies of readership of academic journals are of interest to the different stakeholders involved in scholarly communication: authors, editors, publishers, librarians, etc. Methods to measure journal readership have traditionally ranged from participative methods such as surveys and interviews, to unobtrusive methods such as shelving statistics, document delivery data, download figures and citation counts. Recently, altmetric indicators have been proposed as a complement to citation metrics for research evaluation since they are available sooner and can capture impact beyond the academic community. In a similar fashion, altmetric indicators such as bookmarks in reference management software might be used to supplement other methods for estimating journal readership.

The results of this study confirm the viability of using bookmark counts to explore the behaviour of users of scholarly literature. The coverage of LIS literature by Mendeley is very extensive. By the first quarter of 2015, Mendeley covered 61% of the LIS literature indexed in the Social Sciences Citation Index in the previous 20 years. This is higher than the percentage observed in previous research, which situated Mendeley's coverage of the Social Sciences literature somewhere between 47% and 58% (Mohammadi and Thelwall, 2014; Mohammadi et al., 2015). Given that our research analysed a longer period of time than previous studies focusing on more recent literature, the higher coverage observed in this study is particularly significant. Actually, the Mendeley coverage for LIS papers published in the last decade (2005-2014) was higher than 70%. In a similar fashion, although the percentage of LIS articles cited in Web of Science is higher than the percentage of LIS articles bookmarked in Mendeley for papers published before 2009, the situation is reversed for papers published since 2010. This result confirms that altmetrics offers the important advantage of speed compared to citations for estimating readership.

In addition to speed, another possible advantage of altmetric indicators over citation counts is that they can help to estimate impact outside the academic community. This is especially interesting in fields such as LIS where practicing professionals account for nearly a quarter of the literature published worldwide (Walters and Wilders, 2016). Our results show that, for LIS scholarly literature, librarians and other professionals account

for more than 10% of the bookmarks, even though they still come behind PhD and postgraduate students. The number of librarians who use the tool to manage professional literature may also help to explain the higher coverage of LIS literature in Mendeley compared to other disciplines in other Social Sciences, since in other disciplines most readers are academics – mainly students but also faculty and researchers (Mohammadi et al., 2015).

The distribution of both bookmarks and citations among papers is skewed, but the rankings of papers by number of bookmarks and by number of citations are only partly coincidental. That means that there are articles which have a high number of bookmarks but are not among the top papers by number of citations and vice versa. The reasons for these discrepancies may include (in addition to the different life-cycles of bookmarks and citations) the presence of articles that attract communities which use research but do not cite it, such as students and professionals (Thelwall, 2016).

The results also show that the share of reviews bookmarked in Mendeley is higher than that of articles. Additionally, the average and median number of readers per document is higher for reviews than for articles. This is consistent with previous studies that show that literature reviews are more often cited than regular articles, most likely due to their breadth (Teixeira et al., 2013).

English is the lingua franca of science, and LIS is no exception to this rule. More than 95% of the LIS literature indexed in the Social Sciences Citation Index is in English. In terms of readership, papers in English are more often bookmarked in Mendeley than papers in any other language. Additionally, there are clear differences in the bookmarking of LIS journals. The share of bookmarked articles varies dramatically from journal to journal: some journals have no articles bookmarked at all, while others have nearly all articles bookmarked by at least one reader.

3.4. Conclusions

The results of this study show that altmetrics can be used as a source of data in information behaviour studies. Reference management software provides an unobtrusive means of capturing reading habits in scholarly literature that are useful to all the stakeholders in the scholarly communication system.

The application of altmetric indicators to supplement citations counts in order to estimate readership presents two advantages over the use of citations alone. Bookmarks are available sooner, as shown by the fact that the percentage of recent literature bookmarked in Mendeley is much higher than the share of literature cited. Additionally, bookmarks are useful to capture usage beyond the academic community, since reference management software can be employed by professionals to manage the literature. This is especially relevant in fields such as LIS where practicing professionals account for a large part of the use of the literature.

Chapter 4. A longitudinal study of the bookmarking of library and information science literature in Mendeley (Study 2)

Chapter 4. A longitudinal study of the bookmarking of library and information science literature in Mendeley

Speed and breadth have been suggested as two advantages of altmetrics over citation counts since they might estimate impact immediately after publication and beyond the academic community of authors. As mentioned in Chapter 2, in order to investigate the validity of these claims, the purpose of this study is to compare the evolution in the coverage of LIS literature in altmetric sources with the evolution in the coverage of the same literature in citation indexes. We performed a fifteen-month longitudinal study of the evolution of bookmarks in Mendeley for a set of 3813 articles published in Library and Information Science in 2014. In this Chapter, you can find methods used for data collection and data analysis, sources used, results, conclusion and discussion of Study 2.

4.1. Methods and data collection

According to Moed (2005), longitudinal studies can be conducted using either a synchronous or a diachronous approach.In the former, the researcher analyses the number of citations or downloads (in our case, bookmarks) as a function of the papers' publication date, which is fixed. In the latter, the researcher measures the citations or downloads (or bookmarks) for a set of papers as a function of time. In order to conduct a synchronous analysis, the number of bookmarks for a set of papers published in a given year were recorded each month for fifteen months. In February 2015, we used the Social Sciences Citation Index to retrieve a total of 3813 papers published in 2014 in the category "Information Science & Library Sciences". The search was limited to articles (3750 papers) and reviews (63 papers), and our analysis excluded all other types of document, such as book reviews, editorials, letters or proceedings. Throughout this article we use the term "papers" to refer only to this set of articles and reviews. From March 2015 to May 2016, each paper was searched monthly in Mendeley using Webometric Analyst software(Thelwall, 2009) using a query containing the title of the article, the first author's last name, the year of publication and the Digital Object Identifier (DOI). For each paper, the number of users, their academic status and their country of residence were recorded from Mendeley on the first Thursday of each month. At this point, it should be borne in mind that Mendeley only returns the three main academic status categories and the three main countries of residence for the users of each paper. Mendeley's fifteen academic

status categories were merged into six broader categories in order to facilitate the analysis (Table 8).

Table 8. Mendeley user categories by academic status

Mendeley categories	**Merged**
PhD student	PhD student
Doctoral student	
Postdoc	Postdoc
Student bachelor	Student
Student master	
Postgraduate	
Professor	Faculty
Senior lecturer	
Assistant professor	
Associate professor	
Lecturer	
Researcher academic	Researcher
Researcher non academic	
Librarian	Professional
Other professional	

The original categories provided by Mendeley overlapped and were difficult to distinguish, since users are obliged to select one category but are not provided with thorough definitions. Actually, since the study was conducted, Mendeley itself has grouped some categories, so the options offered now to a new user of the platform do not correspond to those available at the time of data collection. The breakdown of users by academic status in the last month considered, May 2016,seemed to be erroneous, since the results were identical to those in April 2016. Therefore, the results for May 2016 were removed from the analysis of users' academic status. Regarding the users' country of residence, the analysis compared the United States (US), the 28 member states of the European Union (EU-28) and 90 countries with a gross domestic product per capita based on purchasing power parity (PPP GDP) below 25,000 current international dollars (World Bank, 2016).The papers' publication date was set as 2014. While this was the year of publication of the papers' version of record, some of the earliest articles may already have been published prior to the end of 2013. Similarly, some articles dated as 2014 could have been published in early 2015. This is a limitation for a longitudinal study of a fast-changing feature like bookmarks.

4.2. Data analysis and Results

4.2.1. Evolution in the coverage of LIS literature in Mendeley

Out of the 3,813 LIS papers published in 2014, 2,342 (61.4%) had been saved in Mendeley by at least one user by the beginning of March 2015 (Figure 4). Fifteen months later, in May 2016, the coverage had risen to 87.6%, i.e., 3,341 papers had been bookmarked at least once.

While Figure 4 shows two peaks in the share of articles bookmarked—a soft one in September 2015 and a sharper one in January 2016—it is unclear whether these increases reflect a change in users' bookmarking behaviour or are caused by changes in the software used for data collection, which was upgraded several times during the study period. In particular, a few odd cases were observed in the monthly evolution in the number of bookmarks. Despite an increase in the number of bookmarks and the percentage of articles bookmarked from August to September 2015, 179 papers that had 10 or more bookmarks experienced a decrease between 50% and 100% in the number of bookmarks over that period. Most of these papers fell to zero bookmarks and then recovered them in January 2016, a situation that helps to explain the increase observed in the latter month. Two additional situations were observed in which papers with more than 10 bookmarks lost more than 50% of them from one month to the next, affecting 21 and 26 papers, respectively.

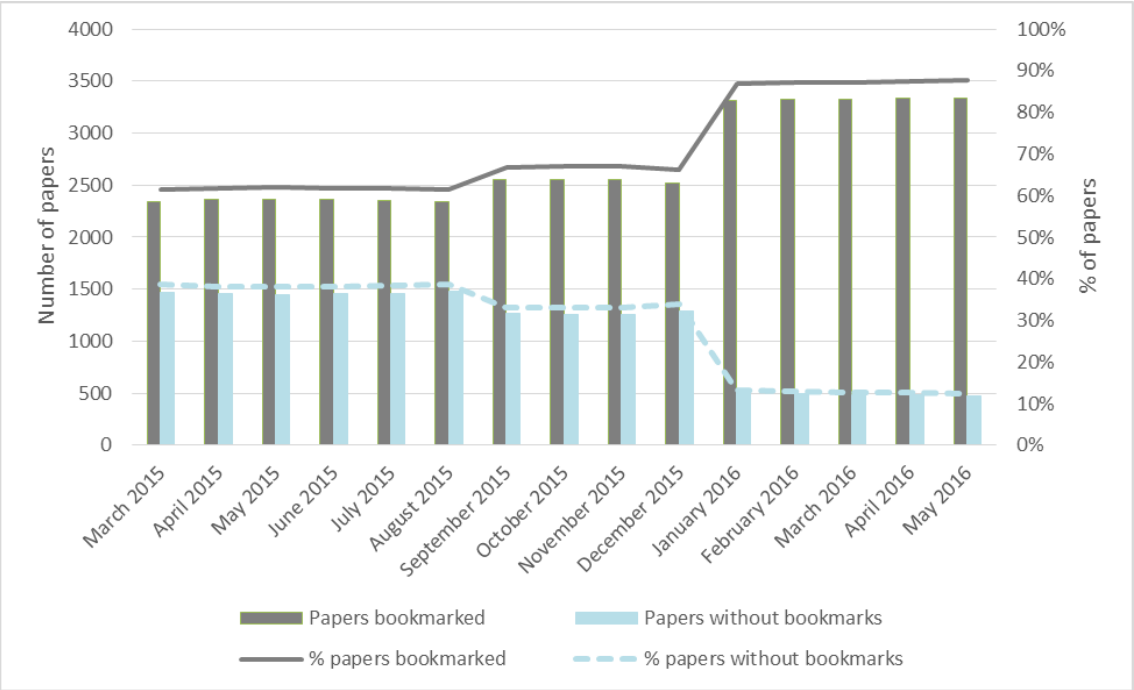

Figure 4. Evolution in the coverage of LIS literature bookmarked in Mendeley

A potential advantage of bookmarks over citations is speed, since articles can be bookmarked immediately after publication, whereas citations need much more time to accrue. Figure 5 shows that by March 2015, 61.4% of the LIS papers published in 2014 had been bookmarked at least once in Mendeley, whereas only 16.4% had been cited in WoS. At the end of the study period, in May 2016, 87.6% of the LIS papers published in 2014 had been bookmarked, whereas only 55.0% had received at least one citation.

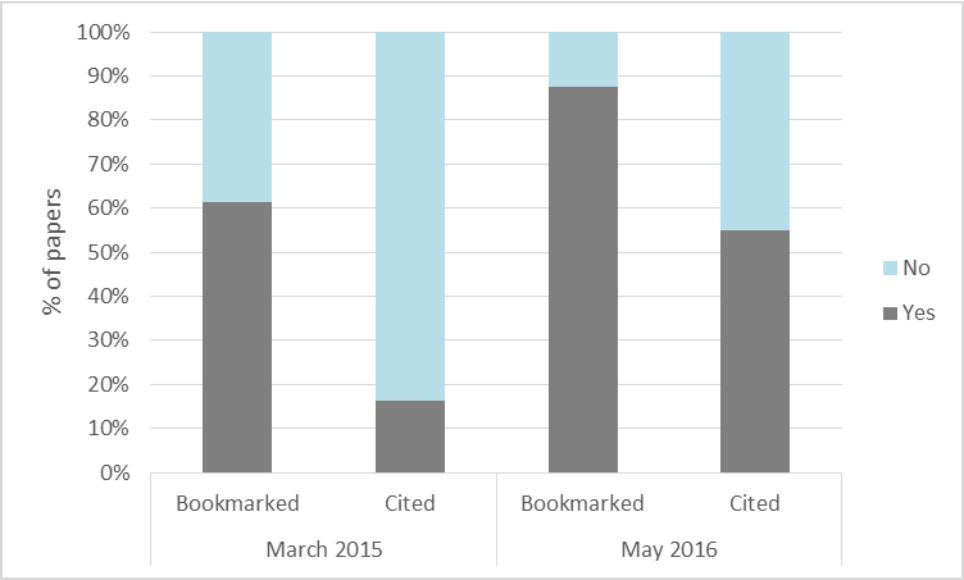

Figure 5. LIS literature bookmarked and cited

Figure 6 supplements the synchronous approach with a diachronous analysis of the number of bookmarks and citations in May 2015 for a set of 54,665 LIS papers published between 1995 and 2014. The results show that the percentage of cited literature is higher for papers published up to 2009, i.e., six years prior to data collection. For recent papers, i.e., those published in the last six years, the share of literature bookmarked is higher than that of literature cited. Although the shares of literature bookmarked and cited both decline for recent papers, the decline is much more acute for citations than for bookmarks. When interpreting these results, it must be borne in mind that the first public beta version of Mendeley was released in August 2008.

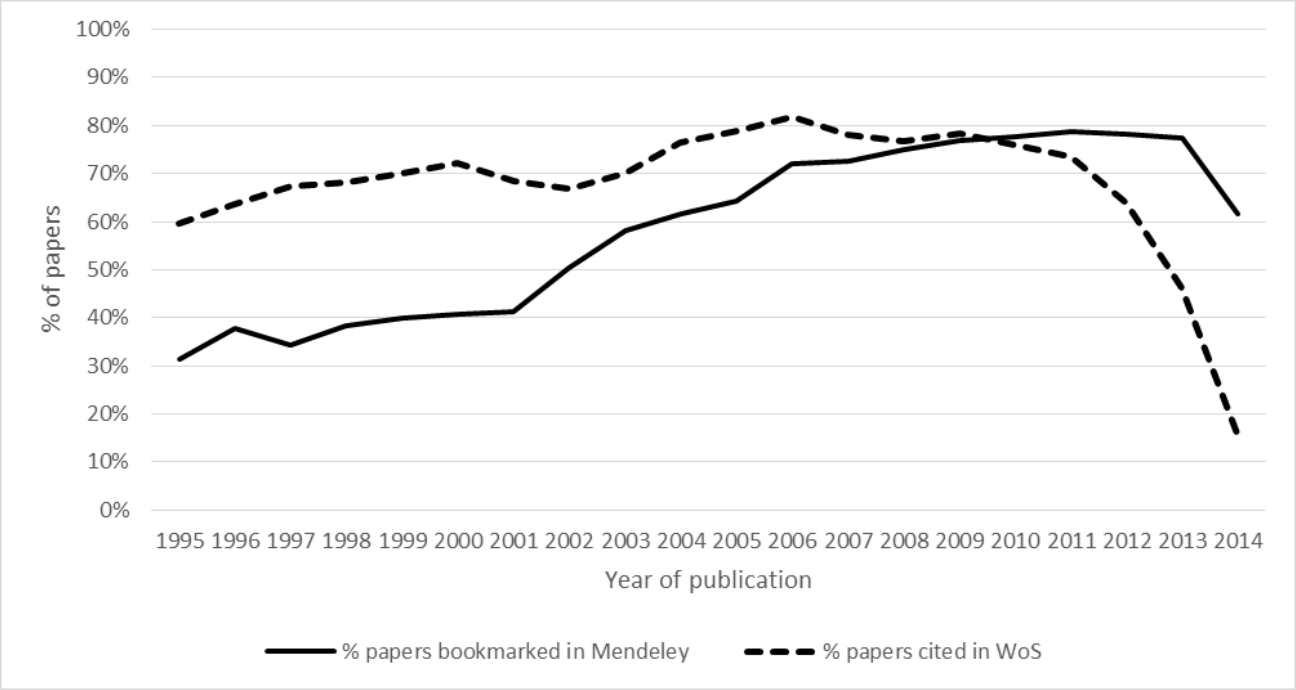

Figure 6. Evolution in the coverage of LIS literature bookmarked in Mendeley and cited in WoS by year of publication (n=54,655 articles, data collected in March 2015)

Table 9 details the bookmarking data for the set of papers. In the fifteen-month study period the number of bookmarks multiplied fourfold, from 19,563 to 83,845, whereas citations multiplied by six, from 993 to 6,515. Not only did the median and the average number of bookmarks per paper increase, but so did the scattering of the bookmarks across the dataset. Thus, as Figure 7 shows, the top 20% of the most frequently bookmarked articles in March 2015 accounted for 70% of the bookmarks. Fifteen months later, the set of 20% most frequently bookmarked articles accounted for 60% of the bookmarks.

Table 9. Evolution in readership of LIS papers in Mendeley

	3/15	4/15	5/15	6/15	7/15	8/15	9/15	10/15	11/15	12/15	1/16	2/16	3/16	4/16	5/16
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q1	0	0	0	0	0	0	0	0	0	0	4	4	4	4	5
Med	2	2	2	3	3	4	6	6	7	7	11	11	11	12	13
Q3	7	8	9	10	11	14	15	16	18	19	24	25	25	26	28
Max	94	118	129	145	153	164	207	215	234	262	245	247	253	268	291
Average	5.3	5.9	6.7	7.6	8.1	10.1	12.0	12.5	13.5	14.6	18.6	19.0	19.4	20.6	22.0
SD	8.9	10.1	11.3	12.4	13.2	16.4	18.3	18.9	20.3	21.9	24.2	24.7	25.5	27.1	28.7
Bookmarks	19563	22061	25110	28166	30203	37573	44547	46457	50090	54274	69395	71112	73816	78621	83845
Citations	993	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	6515

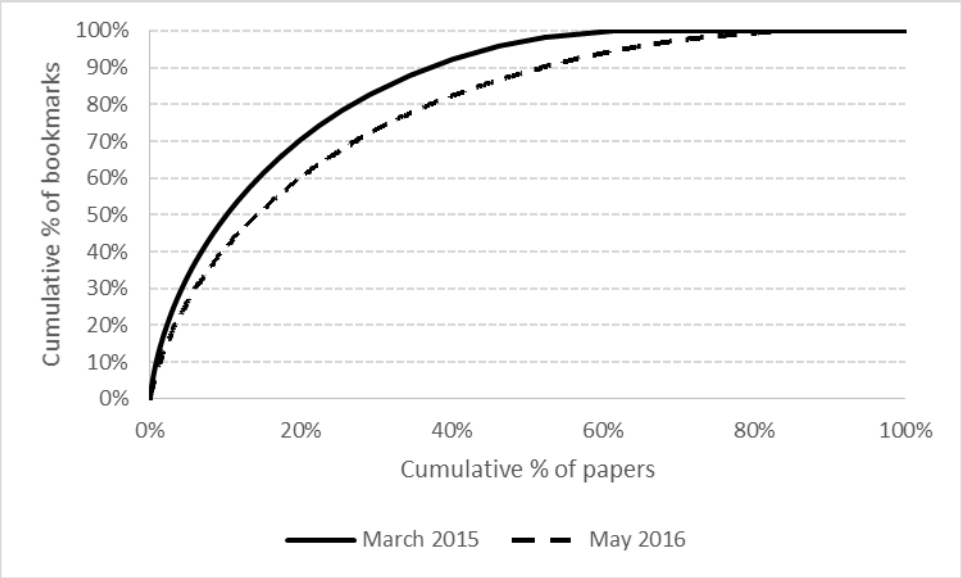

Figure 7. Scattering of bookmarks across the LIS literature

The correlation between bookmarks and citations was moderate throughout the study period, rising slightly from Spearman's rho=0.52 in March 2015 to 0.56 in May 2016. When comparing the rankings of top papers by number of bookmarks and number of citations, we observe significant differences. Specifically, in March 2015 there were 94 papers in the top 20% of articles by bookmarks (they had 30 or more bookmarks each). However, only 15 of these papers (16%) were simultaneously in the top 20% of articles by number of citations (they had received four or more citations each). By May 2016 the overlap had increased to 30%. That is, 41 papers out of the 128 in the top 20% of papers

ranked by bookmarks (93 or more bookmarks each) were also in the top 20% of papers by citations (i.e., they had received ten or more citations each).

4.2.2. Bookmarks by users' academic status

Fig. 8 shows the evolution in the share of bookmarks by users' academic status. The share of papers bookmarked by faculty increased sharply during the study period, rising from 13% of the bookmarks in March 2015 to 38% in the last few months analysed. The opposite trend is observed among students, who were responsible for 30.7% of the bookmarks in March 2015, a percentage that fell to 14.3% at the end of the study period. Similarly, the share of bookmarks by professionals and researchers gradually declined, whereas that of PhD students remained fairly stable at around 30% throughout the whole study period.

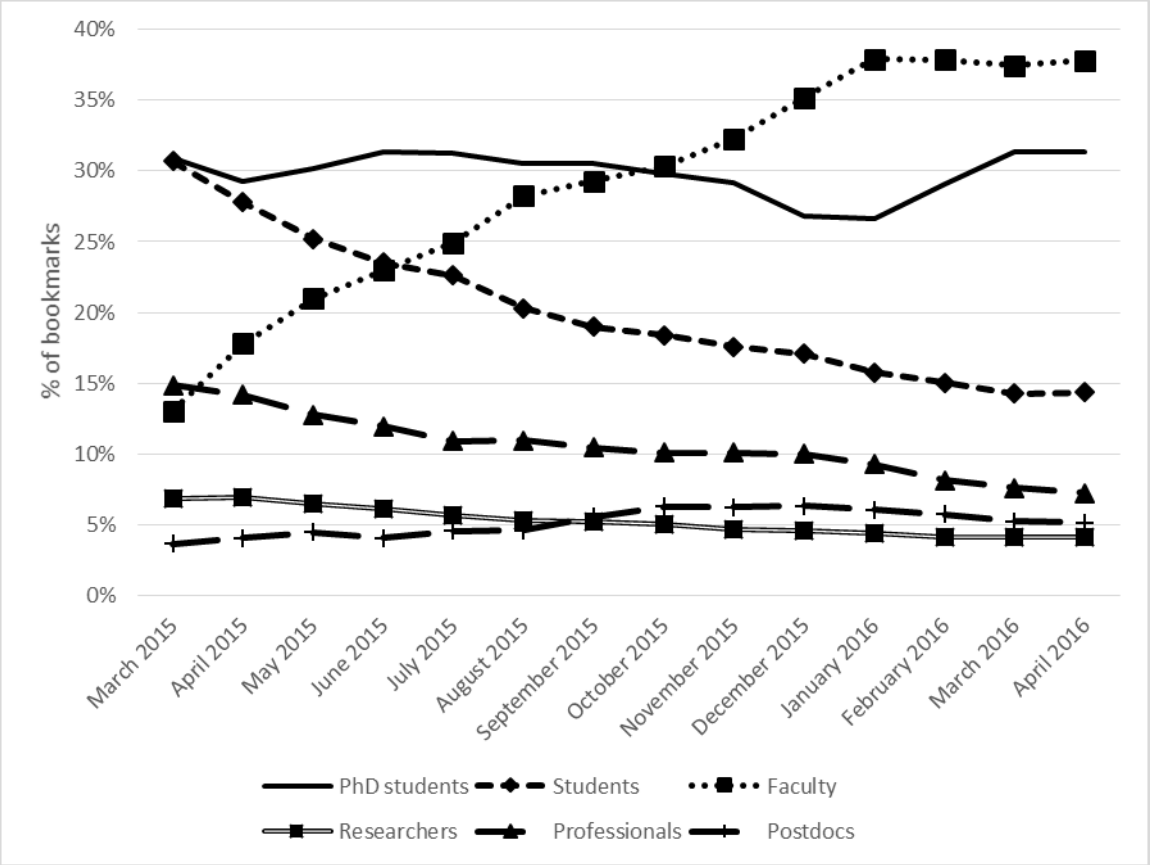

Figure 8. Evolution in the share of bookmarks by users' academic status

4.2.3. Bookmarks by users' country of residence

LIS articles were bookmarked by users based in 165 countries. In order to facilitate the analysis, Fig. 9 presents the evolution in the share of bookmarks made by users in the United States (US), in the 28 member states of the European Union (EU-28) and in 90 countries with a gross domestic product per capita based on purchasing power parity below 25,000current international dollars (PPP GDP < $25,000). All together, these three categories of country represent approximately80% of the bookmarks in the data set. Results: show that the share of bookmarks by US users fell from 22.5% at the beginning of the study period to 19.0% at the end, whereas the share in countries with a lower GDP increased from 17.8% to 24.6%. The percentage of bookmarks made by users based in Europe remained stable at around 40% throughout the whole study period.

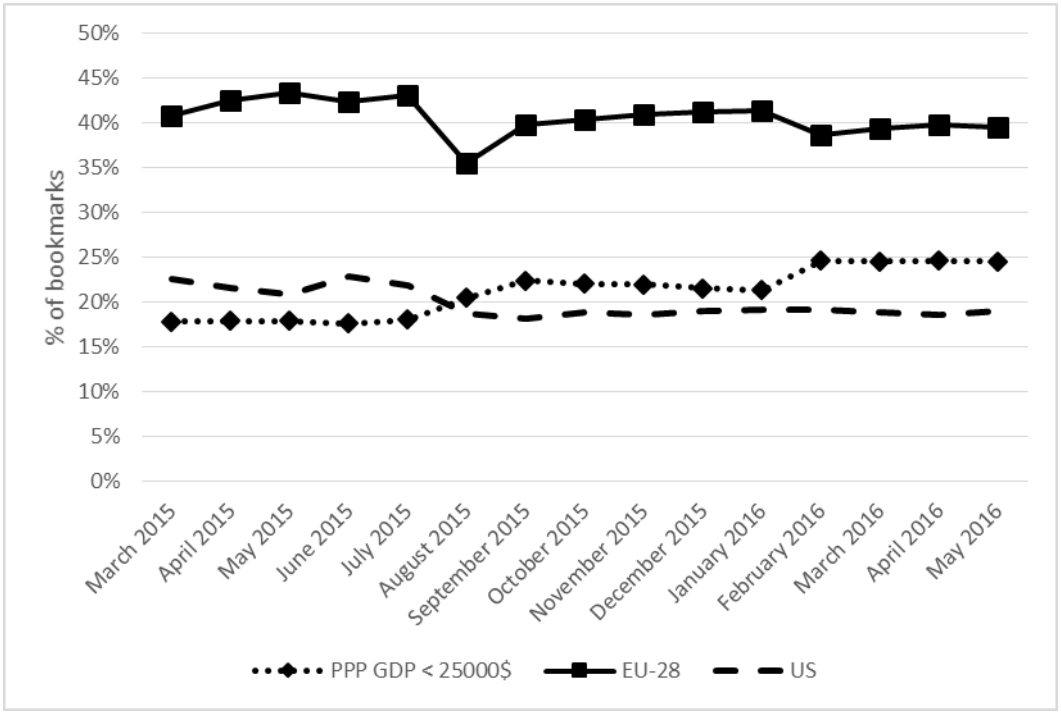

Figure 9. Evolution in the share of bookmarks by users' country of residence

4.2.4. Bookmarks of articles and reviews

Figure 10 shows that reviews were bookmarked more frequently than articles. At the beginning of the study period, in March 2015, 61.2% of the articles had been bookmarked at least once in Mendeley compared to 76.2% of the reviews. Fifteen months later, the percentages had risen to 87.5% of the articles and 96.8% of the reviews. The median number of bookmarks per review was about 2.5 times higher than the median number of bookmarks per article throughout the whole study period.

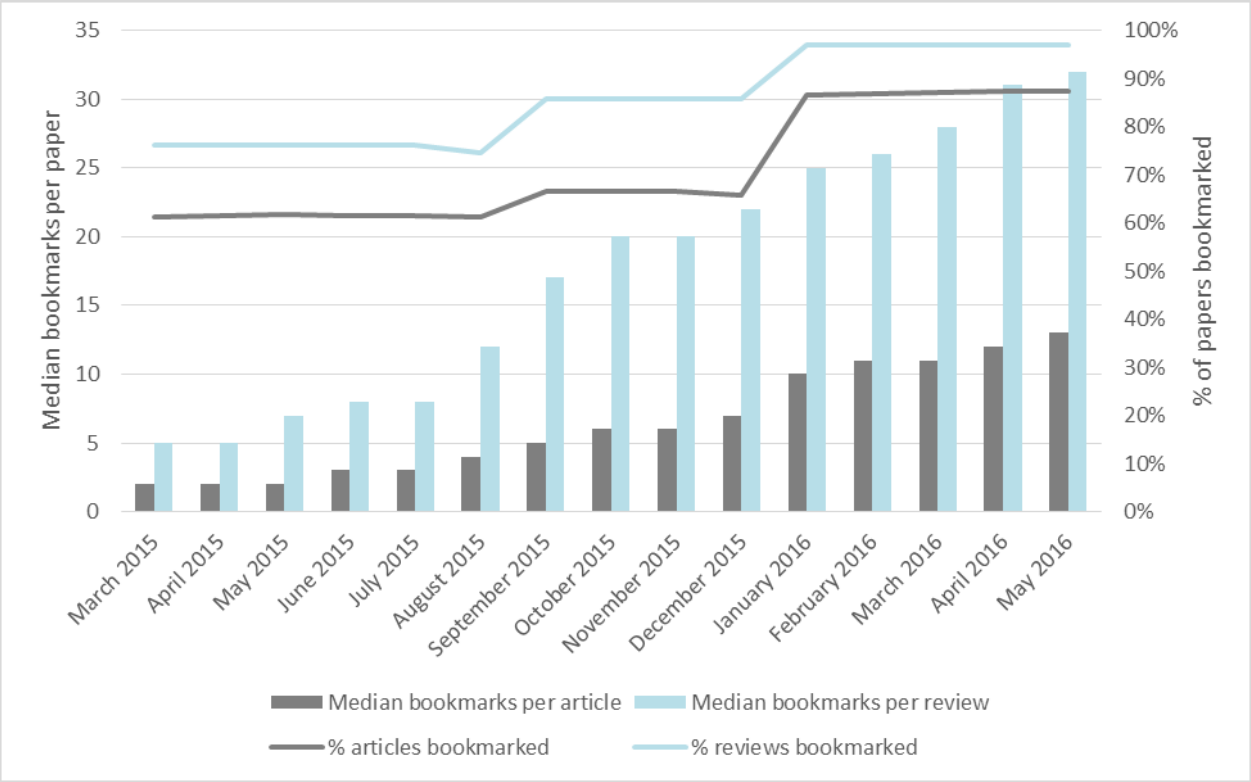

Figure 10. Evolution of the bookmarks of articles and reviews

4.3 Discussion and conclusions

This study provides evidence to support the validity of two suggested advantages of bookmarks over citations in estimating the impact of scientific literature. Papers are bookmarked immediately after publication, whereas citations need much more time to accrue. Thus, nearly nine of every ten LIS papers are bookmarked in Mendeley in the year following publication, a figure that greatly exceeds that of articles cited. This result is consistent with results observed in the only two previous studies to investigate the role

of time in the evolution of Mendeley bookmarks. Specifically, Maflahi and Thelwall (2016), using a diachronous approach, observed that, for a set of four LIS journals, papers initially attracted more Mendeley bookmarks than Scopus citations, but the situation reversed after approximately seven years. A further study expanded the analysis to five disciplines and confirmed that the correlation between bookmarks and citations tended to increase over five years and then stabilise (Thelwall and Sud, 2016).

Similarly, the results are consistent with those observed in studies on the usage of scholarly literature, such as those by Tenopir and King (2000), who showed that nearly two-thirds of all use occurs in an article's first year of publication. Nicholas et al. (2005) reached a similar conclusion by analysing download data that showed that about 55% of use is for items published within the previous 15 months. Schlögl et al. (2014) also showed that the highest number of downloads of papers are usually made in the publication year and immediately afterwards. All in all, these results suggest that Mendeley bookmarks can be a good complement to citations in the assessment of young articles, provided they are in a discipline with a high level of Mendeley users. According to our results, this is the case of LIS where Mendeley coverage of the literature is higher than in other Social Sciences where the coverage is somewhere between 47% and 58% (Mohammadi and Thelwall, 2014; Mohammadi et al., 2015).

The core sets of papers by number of bookmarks and number of citations only partly coincide. This means that some articles having a high number of bookmarks are not among the top papers by number of citations and vice versa. In addition to the different life-cycles of bookmarks and citations, the reasons for this discrepancy may include the presence of articles that attract communities which use research but do not cite it, such as students and professionals (Thelwall, 2016). This evidence supports the second potential advantage of bookmarks over citations, i.e., that bookmarks reflect impact beyond the academic community of authors. Our results show that a significant share of the bookmarks of LIS literature are made by students and professionals who are not necessarily authors and therefore do not cite. At this stage some unexpected results were observed in the evolution in the share of papers bookmarked by the different categories of users. While the amount of bookmarks increased across all categories of users, the increase in the share of bookmarks made by faculty was significantly higher. This suggests that faculty do not bookmark papers immediately after publication, but do tend

to bookmark papers over time. Another explanation might be an increase in the population of faculty users who employ Mendeley. Further research will be necessary to interpret this result. A tentative analysis was performed in order to compare the journals bookmarked by different categories of users in terms of their orientation (computer science, librarianship, management, etc.) but no significant results were obtained and the analysis was discarded.

Access to scholarly information is a major concern in developing countries where fewer resources are available for journal subscriptions. The high cost of Western scientific journals poses a major barrier to researchers in developing nations (Davies and Walters, 2011). This situation may be mirrored in the evolution in the share of bookmarks by users' country of residence. Thus, a reduction in the share of articles bookmarked by users based in the US is compensated by an increase in the share of bookmarks in less wealthy nations with a lower GDP. With the passage of time, open-access versions of articles become available in repositories and other websites that may stimulate an increase in the amount of bookmarks. Actually, the differences in the evolution of the shares of bookmarks by country of residence could be sharper than suggested by the results, since Mendeley only returns the three main countries of residence for the users of each paper.

The results also show that the share of reviews bookmarked in Mendeley is higher than that of articles. Additionally, the average and median number of users per document is higher for reviews than for articles. This is consistent with previous studies that show that literature reviews are more often cited than regular articles, most likely due to their breadth (Teixeira et al., 2013; Schlögl et al. 2014).

In addition to the fact that Mendeley only returns the three main academic status categories and the three main countries of residence for the users of each paper, two further limitations need to be borne in mind when interpreting the results. First, upgrades in the software employed for data collection may have affected the reliability of the data as reflected by fluctuations in the monthly amount of bookmarks for some papers. Second, the papers' publication date was set as 2014, but this is only the year of publication of the version of record and their dates of online availability might well precede or extend beyond that year.

Chapter 5. Methodological issues in measuring citations in Wikipedia: a case study in Library and Information Science (Study 3)

Chapter 5. Methodological issues in measuring citations in Wikipedia: a case study in Library and Information Science

Wikipedia citations have been suggested as a metric that partially captures the impact of research, providing an indication of the transfer of scholarly outputs to a wider audience beyond the academic community. As mentioned in Chapter 2, the purpose of this study is to explore the coverage of Library and Information Science (LIS) literature in Wikipedia. We explore the coverage of Library and Information Science literature published between 2001 and 2010 in Wikipedia, paying special attention to the methodological issues involved in counting Wikipedia citations. In this Chapter, you can find methods used for data collection and data analysis, sources used, results, conclusion and discussion of Study 3.

5.1. Methods and data collection

In order to conduct the study, we retrieved the 26,542 articles and reviews indexed in the category "Information Science & Library Sciences" of the *Social Sciences Citation Index* in the *Web of Science* published between 2001 and 2010.

Afterwards, we searched each of these articles in Wikipedia and retrieved all the entries where they were cited. In order to do so, the advanced search feature of Google was employed, searching all the words in the article title as an exact phrase and narrowing the results to those in the domain "wikipedia.org". In the case of articles with very short titles (three or four words), the name of the first author was added to the query and results were checked manually. All the searches were conducted between the second half of 2016 and early 2017 to allow for an extended period of at least five years since the publication of an article. Citation analysis studies usually employ a shorter citation window (impact factors, for instance, are based on the citations received by articles published in the previous two years). However, since this study focuses on citations outside the academic community, an extended citation window seemed appropriate.

Any citation to an article in a Wikipedia entry was recorded, either in the "references" section, as an "additional reading" or embedded in the text (for instance, in a section of a Wikipedia entry entitled "Example studies that have leveraged the IS success model"). In

the case of articles cited in several Wikipedia entries, all the instances were recorded. Similarly, the citation of the article in different language versions of a single Wikipedia entry was recorded. At this point, it should borne in mind that the different language versions of a single Wikipedia entry are not translations but are edited independently and therefore can cite different sources.

Finally, when the reference included a link to an external source, we visited the website to know whether the full-text of the article was available in OA. Again, in the case of articles cited in several Wikipedia entries, all the references were checked since they may link to different sources. However, in the case of citations to a single article in different language versions of the same Wikipedia entry, only the reference in the English version (or the first retrieved version if the article was not cited in the English version) was checked.

5.2. Data analysis and Results

5.2.1. Limitations for counting Wikipedia citations

The retrieval of the Wikipedia citations to the academic articles in the sample proved to be a difficult task due to the lack of standardization of the bibliographic references. Table 10 provides examples of the most frequently observed problems. Thus, example 1 illustrates the case of a reference that just includes the article's title linking to the full-text stored in the publisher's website. Reference 2 shows a slightly more complete citation including, in addition to the article title and URL pointing to the publisher, the journal and year of publication. Meanwhile, reference 3 includes the author's name, year of publication, title, and URL. In this case, the link points to a postprint copy of the article deposited in an institutional repository.

Table 10. Examples of incomplete references in Wikipedia

1. Reference including title and URL
• Interpolation of the extended Boolean retrieval model Source: https://en.wikipedia.org/wiki/Extended_Boolean_model
2. Reference including title, URL, journal and year
• "But the data is already public": on the ethics of research in Facebook, *Ethics and Information Technology*, 2010 Source: https://en.wikipedia.org/wiki/Michael_Zimmer_(academic)
3. Reference including author, year, title and URL
14. ^ Gorman, M. (2001). "Values for Human-to-Human Reference". p179 Source: https://en.wikipedia.org/wiki/Michael_Gorman_(librarian)
4. Reference including authors, year and title
• Walther, Joseph B and D'Addario, Kyle P (2001). *The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication* Source: https://en.wikipedia.org/wiki/Social_information_processing_(theory)
5. Reference including authors, year, title, journal, volumen, issue, pages and DOI
• Walther, Joseph B.; Kyle P. D'Addario (2001). "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication". *Social Science Computer Review*. **19** (3): 324–347. doi:10.1177/089443930101900307. Source: https://en.wikipedia.org/wiki/Joseph_Walther
6. Reference including abbreviated journal title
^ a b Kim, Matthew; Johnson, Kevin (2002). "Personal health records: evaluation of functionality and utility". *JAMIA*. (2): 171–180. doi:10.1197/jamia.M0978. PMC 344574. PMID 11861632. Source: https://en.wikipedia.org/wiki/Personal_health_record
7. Erroneous reference inverting authors' names and surnames (autors are Maryam Alavi, George M. Marakas and Youngjin Yoo):
7. ^ 7.0 7.1 Maryam A.&George M.&Youngjin Y.(2002). A Comparative Study of Distributed Learning Environments on Learning Outcomes. Information Systems Research, Vol. 13, No. 4 Source: https://en.wikipedia.org/wiki/Social_learning_theory (Hebrew version)

The degree of completeness of the references varies from entry to entry, even for a single article. Example 4 shows a reference just including the names of the authors, year of publication and title, whereas reference 5 provides a much more detailed citation for the same article obtained from a different Wikipedia entry. Even in the case of references relatively complete, abbreviated journal titles can make them difficult to retrieve, as in example 6 where the *Journal of the American Medical Informatics Association* has been

abbreviated to *JAMIA*. Some references contained errors, such as that in example 7 where authors' names and surnames had been inverted.

The use of the "cite journal" template [5] to create citations for scientific papers is inconsistent. It is frequent to find Wikipedia entries where "references" have been edited using the recommended template, but citations included in sections such as "further reading", "select bibliography" or "external links" have not. This is the case, for instance, of examples 1 and 2 in Table 1. Even when the citation template is used, examples in Table 1 show that multiple parameters may be missing. The inclusion of a DOI in the reference could be used to automatically extract Wikipedia citations to academic articles. However, for articles published in 2010, the latest year considered in our study, just 61 references out of 115 (53%) included a DOI. Again, it is possible to find examples of a single article cited in several Wikipedia entries with and without a DOI.

5.2.2. Proportion of LIS literature cited in Wikipedia

Overall, just 2.9% (766 articles) of the LIS output published between 2001 and 2010 and indexed in the *Social Sciences Citation Index* had been cited in Wikipedia by the time of data collection. Since some articles had been cited in several Wikipedia entries, the total amount of citations retrieved rose to 982 (Table 11).

Table 11. LIS literature cited in Wikipedia by publication year

Year	Articles in WoS	Articles cited in Wikipedia	Citations of LIS literature	Wikipedia citations in authors' biographies	OA citations
2001	2,349	51 (2.2%)	77	17 (22.1%)	14 (18.2%)
2002	2,337	59 (2.5%)	66	13 (19.7%)	16 (24.2%)
2003	2,315	62 (2.7%)	73	17 (23.3%)	16 (21.9%)
2004	2,173	80 (3.7%)	111	18 (16.2%)	36 (32.4%)
2005	2,499	84 (3.4%)	105	21 (20.0%)	27 (25.7%)
2006	2,512	79 (3.1%)	99	12 (12.1%)	30 (30.3%)
2007	2,820	88 (3.1%)	122	6 (4.9%)	45 (36.9%)
2008	2,962	94 (3.2%)	115	13 (11.3%)	37 (32.2%)
2009	3,165	86 (2.7%)	99	9 (9.1%)	36 (36.4%)
2010	3,272	83 (2.4%)	115	7 (6.1%)	49 (42.6%)
Total	**26,542**	**766 (2.9%)**	**982**	**133 (13.5%)**	**306 (31.2%)**

5.2.3. Citations in Wikipedia biographies

As could be expected, Wikipedia entries citing LIS literature were related to topics in the field. Frequently, these Wikipedia entries were biographical articles about well-known scholars in the field (such as Marcia J. Bates, Hope A. Olson or Tom Wilson, to name but a few) accounting for their education, work, awards, etc.

Most of these biographical entries include a list of publications authored by the scholar in question (see, for instance, example 5 in Table 10). At this point, it is interesting to note that 13.5% of the Wikipedia citations retrieved in our study were made in biographical entries about one of the authors of the article cited (Table 11). The amount of citations in authors' biographical entries was especially significant for articles published in the initial five years covered in our study, although it decreased for more recent literature.

5.2.4. Open access availability of articles cited in Wikipedia

Scholarly journals often require expensive subscriptions. It is therefore questionable whether Wikipedia contributors may have access to these sources or whether they rely on

OA sources to edit entries. Our results show that 31.2% of the Wikipedia citations linked to an OA source, with this percentage increasing for more recent literature.

At this point, we counted separately the citations to a single article in several Wikipedia entries, since a reference may link to an OA source in one entry but not in another. For instance, Table 12 shows four different linking options for a single article in four Wikipedia entries: reference 1 does not include any link; reference 2 is linked to the publisher's website that requires a subscription to gain access to the full-text; reference 3 includes a broken link to one of the co-author's personal website; finally, reference 4 links

to a freely available post-print copy of the article stored in the Internet Archive saved version of the page linked in reference 3.

Table 12. Examples of different linking options to a single article

1. Reference without link
9. ^ Sondergaard T. F.; Andersen J.; Hjorland B. Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. Journal of Documentatio 2003, 59, (3), 278–320. Source: https://en.wikipedia.org/wiki/Grey_literature
2. Reference with link to publisher
2. ^ Søndergaard, T. F.; Andersen, J.; Hjørland, B. (2003). "Documents and the communication o scientific and scholarly information: Revising and updating the UNISIST model". *Journal of Documentation*. **59** (3): 278. doi:10.1108/00220410310472509. Source: https://en.wikipedia.org/wiki/UNISIST_model
3. Reference with broken link to co-author's personal website
Fjordback Søndergaard, T.; Andersen, J. & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. *Journal of Documentation, 59*(3), s. 278-320. http://www.db.dk/bh/UNISIST.pdf Source: https://en.wikipedia.org/wiki/Source_literature
4. Reference with link to copy of co-author's personal website in Internet Archive
• Fjordback Søndergaard, Trine; Andersen, Jack & Hjørland, Birger (2003). Documents an the communication of scientific and scholarly information. Revising and updating the UNISIST model. Journal of Documentation, 59(3), 278-320. (Available at: http://web.archive.org/web/20050320083023/http://www.db.dk/bh/UNISIST.pdf) Source: https://en.wikipedia.org/wiki/Scientific_communication

The 306 references including an OA link pointed to three kind of sources in a balanced manner: publishers' websites (fully OA journals, delayed OA articles after an embargo period, OA articles in hybrid journals, etc.): 39.2%; repositories (disciplinary or institutional): 30.4%; and websites (personal, departmental, social networks, etc.): 30.4%.

In the case of citations to articles published in fully OA journals, it could be expected to systematically find links to the full-text available in the publisher's website. However, this was not always the case. Table 13 shows two examples of references to articles

available in OA journals that are not linked from Wikipedia. The first example corresponds to an article published in *College and Research Libraries*, a journal currently available in OA. The reference provides a (broken) link to the social network Academia.edu. A possible reason for this situation is that the reference, according to the retrieved date, was introduced in December 2010, but *College and Research Libraries* did not become OA until the following year. The second example shows a reference to an article published in *Information Research*, a fully OA journal since its creation. The reference, however, does not include any link. Obviously, if publishers' versions freely available are not always linked by Wikipedia contributors, it is highly plausible that copies deposited in repositories or other sources are not linked either, making results in Table 13 an underestimate of the OA availability of cited sources.

Table 13. Examples of references to OA articles in Wikipedia without links to full-text

1. Reference to an article in *College and Research Libraries*
2. ^ Beall, Jeffrey; Kafadar, Karen (2005). "The Proportion of NUC Pre-56 Titles Represented i OCLC WorldCat". *College & Research Libraries*. **66** (5): 431–5. Retrieved 2010-12-23.
Source: https://en.wikipedia.org/wiki/National_Union_Catalog
2. Reference to an article in *Information Research*
5. ^ Shiyong Lu, Dapeng Liu, Farshad Fotouhi, Ming Dong, Robert Reynolds, Anthony Aristar, Martha Ratliff, Geoff Nathan, Joseph Tan, and Ronald Powell, "Language Engineering for the Semantic Web: a Digital Library for Endangered Languages", *Information Research,* 9(3), Apri 2004.
Source: https://en.wikipedia.org/wiki/Language_planning

5.3. Discussion and conclusions

Citations in Wikipedia have been proposed, among other altmetric indicators, as an alternative to traditional impact metrics. Thus, citations of articles in Wikipedia can be seen as a metric that partially captures the societal and educational impact of the article among a wider audience beyond the academic community. The results of this study,

however, reveal severe limitations regarding the use of Wikipedia citations for research evaluation purposes.

The lack of standardization of Wikipedia references makes difficult to measure them with a minimum level of precision. Contrary to bibliographies in academic publications, where references are edited to guarantee that they are correct, Wikipedia citations are frequently incomplete, or even erroneous. Essential fields for the proper identification of articles, such as authors' names or journal titles may be missing, making impossible the retrieval of citations. This feature, combined with the absence of document identifiers, such as DOIs, impedes to rely on automatic extractions of citations. If professionally edited citation indexes, such as Scopus and Web of Science, have been criticized for their inaccuracies that make difficult the retrieval of some documents and distort bibliometric indicators (Franceschini et al. 2015; 2016), it is hard to think on using Wikipedia citations for research evaluation purposes. Bibliometric indicators based on Wikipedia citations will hardly reach the requirements of robustness and replicability necessary to be used in decision-making processes.

The amount of Wikipedia citations is also too small to be used in research evaluation. Less than 3% of the LIS articles published between 2001 and 2010 had been cited in Wikipedia by 2016. This figure results from a detailed search of individual articles including manual checks but, given the lack of standardization and incompleteness of many references, any automatic attempt to retrieve Wikipedia citations would probably lead to a lower figure. Given the scarce amount of information provided in some references, it is also possible that we have missed some citations. Although our study focuses in a small discipline such as LIS, the results are consistent with those obtained by Lin and Fenner (2014) who found that just 4% of PLoS articles had been cited in Wikipedia and Kousha and Thelwall (2015) who reported that only 5% of the articles indexed by Scopus between 2005 and 2012 had been cited in the encyclopedia.

In addition to the low percentage of the scholarly literature cited in Wikipedia, attention must also be paid to the representativeness of these citations. As stated by Thelwall (2016), the use of Wikipedia citations as a proxy for public interest in research articles is limited due to the intermediating role of Wikipedia contributors, with references reflecting the interest of a small number of researchers and amateurs that are enthusiastic

Wikipedia editors rather than the general public. Although our study does not deal with this issue, our results unveil some features that should also be borne in mind when interpreting Wikipedia citation data. This is the case of the relatively large amount of Wikipedia citations retrieved in the biographies of articles' authors. It is frequent that Wikipedia biographies of relevant scholars list their publications, increasing the amount of citations received by well-known scholars in the field. This behaviour results in a phenomenon of accumulated advantage similar to the Matthew effect. Our results show that this phenomenon is more evident for older literature, suggesting that biographical Wikipedia entries are created for more senior scientists.

The relationship between OA availability and Wikipedia citations is also of interest since we can intuitively assume that easy accessibility makes articles more likely to be referenced (Teplitskiy, Lu and Duede, 2017). Our results show that 31.2% of the Wikipedia citations of LIS literature linked to an OA source, with this percentage increasing for more recent literature. However, this is probably an underestimate of OA availability due to the incompleteness of Wikipedia citations, with links to OA sources frequently missing.

Chapter 6. Conclusions

Chapter 6. Conclusions

This chapter aims to answer all specific questions of each paper, then it presents some limitations of the dissertation and also some recommendations for future research.

6.1. Main research questions

1. **What was Mendeley's coverage in 2015 of the LIS literature published in the 20-year period 1995-2014?**

   The result for this question indicated that around 61% of the LIS literature published in the 20-year were coveraged in Mendeley's 2015.

2. **What was the share of the LIS literature bookmarked in Mendeley compared to the share of the same literature cited in Web of Science?**

   The percentage of cited papers is higher among papers published between 1995 and 2004,the share of cited and bookmarked papers becomes progressively balanced for papers published between 2005 and 2009. In the case of papers published in the last five years (2010-2014), 55% had been cited in Web of Science at the time of data collection, in May 2015, whereas 75% had been bookmarked at least once in Mendeley.

3. **Did the number of bookmarks in Mendeley vary according to the type of paper (article or review)?**

   Yes, 67% of the LIS reviews published between 1995 and 2014 were bookmarked in Mendeley, against 61% of the articles. The median and the average number of bookmarks per review were 2.5 times higher than the number of bookmarks per article.

4. **Did the number of bookmarks in Mendeley vary according to the language of the paper?**

   Yes, the number of bookmarks varied according to the language of the paper. Overwhelming majority (95%) of the LIS papers indexed in the Social Sciences

Citation Index between 1995 and 2014 were in English. The second language by number of papers was Spanish. Overall, the percentages of English and Spanish papers bookmarked in Mendeley were similar. The number of LIS papers in other languages indexed in the Social Sciences Citation Index was much smaller and, in all cases, the percentage of papers bookmarked in Mendeley and the average and median number of bookmarks per paper were smaller than among English and Spanish papers.

5. **What is the academic status of Mendeley users of LIS literature?**

The largest group of users of LIS literature in Mendeley were PhD students (34%), followed by postgraduate student (22%), librarians (8%), Assistant professor (7%), Researcher (6%), Associate professor and Student (bachelor) (5%), Professor and Postdoc (4%), Other profesional (3%) and Researcher (at non/academic institution) (2%).

6. **What are the most widely bookmarked journals in LIS according to Mendeley figures?**

The most widely bookmarked journal is *Information Technology for Development*, followed by *Telematics and Informatics, International Journal of Computer/Supported Collaborative Learning*, *Information System Journal*. The full list is in Annex1.

7. **What is the evolution in the coverage of LIS literature in Mendeley compared to the evolution in the number of citations received in WoS?**

By March 2015, 61.4% of the LIS papers published in 2014 had been bookmarked at least once in Mendeley, whereas only 16.4% had been cited in WoS. At the end of the study period, in May 2016, 87.6% of the LIS papers published in 2014 had been bookmarked, whereas only 55.0% had received at least one citation.

8. What is the evolution in the population of users of LIS literature by academic status?

The share of papers bookmarked by faculty increased sharply during the study period, rising from 13% of the bookmarks in March 2015 to 38% in the last few months analysed. The opposite trend is observed among students, who were responsible for 30.7% of the bookmarks in March 2015, a percentage that fell to 14.3% at the end of the study period. Similarly, the share of bookmarks by professionals and researchers gradually declined, whereas that of PhD students remained fairly stable at around 30% throughout the whole study period.

9. What is the evolution in the population of users of LIS literature by country of residence?

LIS articles were bookmarked by users based in 165 countries. In order to facilitate the analysis, we divida ll these countries in three categories United States (US), in the 28 member states of the European Union (EU-28) and in 90 countries with a gross domestic product per capita based on purchasing power parity below 25,000 current international dollars (PPP GDP < $25,000). All together, these three categories of country represent approximately 80% of the bookmarks in the dataset. Results show that the share of bookmarks by US users fell from 22.5% at the beginning of the study period to 19.0% at the end, whereas the share in countries with a lower GDP increased from 17.8% to 24.6%. The percentage of bookmarks made by users based in Europe remained stable at around 40% throughout the whole study period.

10. Are there any differences in the bookmarking of articles and reviews?

That reviews were bookmarked more frequently than articles. At the beginning of the study period, in March 2015, 61.2% of the articles had been bookmarked at least once in Mendeley compared to 76.2% of the reviews. Fifteen months later, the percentages had risen to 87.5% of the articles and 96.8% of the reviews. The

median number of bookmarks per review was about 2.5 times higher than the median number of bookmarks per article throughout the whole study period.

11. What are the methodological limitations of counting Wikipedia citations?

The lack of standardization of Wikipedia references makes it difficult to measure them with a minimum level of precision. Unlike bibliographies in academic publications, where references are edited to ensure that they are correct, Wikipedia citations are frequently incomplete or even erroneous. Essential fields for the proper identification of articles such as authors' names or journal titles may be missing, making it impossible to retrieve citations. This feature, combined with the absence of document identifiers such as DOIs means that we cannot rely on automatic extractions of citations. It is hard to consider using Wikipedia citations for research evaluation purposes. Bibliometric indicators based on Wikipedia citations will be unlikely to reach the requirements of robustness and replicability necessary to be used in decision-making processes.

12. How much is the proportion of LIS literature cited in Wikipedia?

2.9% (766 articles) of the LIS output published between 2001 and 2010 and indexed in the Social Sciences Citation Index had been cited in Wikipedia by the time of data collection. Since some articles had been cited in several Wikipedia entries, the total number of citations retrieved was 982.

13. What are the characteristics of Wikipedia entries that cite LIS literature?

Wikipedia entries citing LIS literature were related to topics in the field. Frequently, these Wikipedia entries were biographical articles about well-known LIS scholars (such as Marcia J. Bates, Hope A. Olson and Tom Wilson, to name but a few) describing their education, work and awards, among other information.

Most of these biographical entries include a list of publications authored by the scholar in question. In fact, a total of 13.5% of the Wikipedia citations retrieved in our study were made in biographical entries about one of the authors of the cited article. The number of citations in authors' biographical entries was especially significant for articles published in the initial five years covered in our study, although it decreased for more recent literature.

14. How much was the OA availability of the LIS articles cited in Wikipedia?

The 306 references that included an OA link pointed to three kinds of sources in a balanced manner: publishers' websites (fully OA journals; articles that were OA after an embargo period and OA articles in hybrid journals, among others): 39.2%; repositories (disciplinary or institutional): 30.4%; and websites (personal, departmental and social networks, among others): 30.4%.

6.2. Limitations

The first limitation of this research is that Mendeley only returns the three main academic status categories and the three main countries of residence for the users of each paper.

The second limitation is regardings upgrades in the software employed for data collection may have affected the reliability of the data as reflected by fluctuations in the monthly amount of bookmarks for some papers.

The third one is related to the papers' publication date which was set as 2014, but this is only the year of publication of the version of record and their dates of online availability might well precede or extend beyond that year.

The fourth limitation which we can consider as a general limitation. Many scholars use other similar tools, such as EndNote, RefWorks and ProCite, to organize their references, or do not use a reference manager at all.

The fifthe limitation which is related to Wikipedia's entries. The lack of standardization of Wikipedia references makes it difficult to measure them with a minimum level of precisión.

The sixth one is that Wikipedia citations are frequently incomplete or even erroneous. Essential fields for the proper identification of articles such as authors' names or journal titles may be missing, making it impossible to retrieve citations. This feature, combined with the absence of document identifiers such as DOIs means that we cannot rely on automatic extractions of citations.

6.3. Future research

This work provided an overview on articles and reviews. From the quantitative point of view, more studies are needed to explore different document types in different social platforms because document type importance varies by discipline. For example, the book is important in social science and humanities disciplines and conference papers are important in engineering.

Additionally, due to the differences in the levels of activities of scholars across different disciplines, future studies need to find a way to suggest normalized altmetrics for academic fields similar to the normalized citation indicators.

New research to find out who are the users of different social platforms and their motivations for using these tools in their scholarly activities can be interesting.
Comparing the coverage of LIS literature in other social web platforms such as

news sites, blogs, microblogs, Twitter, CiteUlike, social bookmarking tools and other online reference managers is also recommended.

References

References

Amara, N., & Landry, R. (2012). Counting citations in the field of business and management: why use Google Scholar rather than the Web of Science. *Scientometrics*, 93(3), 553–581.

Archambault, É., & Larivière, V. (2009). Origins of Measures of Journal Impact: Contingencies and Consequences. *Scientometrics,* 79(3), 635-649.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries,* 3(7), 1-8.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *Proceedings of the 17th International Conference on Science and Technology Indicators*, Montreal, Canada (pp. 98–109).

Bonasio, A. (2013). Mendeley has 2.5 million users!, *Mendeley Blog*, available at: https://blog.mendeley.com/2013/09/03/mendeley-has-2-5-million-users/ (accessed 1 June 2016).

Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Infor- mation Science and Technology*, 36(1), 2–72.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8 (4), 895-903.

Bornmann, L. & Haunschild, R. (2015), Which people use which scientific papers? An evaluation of data from F1000 and Mendeley. *Journal of Informetrics*, (9) 3, 477-487.

Borrego, A. & Fry, J. (2012), Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy, *Journal of Information Science*, 38(3) , 297–308.

Bosman, J., Van Mourik, I., Rasch, M., Sieverts, E., & Verhoeff, H. (2006). Scopus reviewed and compared. *Utrecht University Library*. Retrieved from http://igitur-archive.library.uu.nl/search/search.php.

Davies, P. M. & Walters, W. H. (2011). The impact of free access to the scientific literature: a review of recent research. *Journal of the Medical Library Association*, 99(3), 208–217.

Evans, P. & Krauthammer, M. (2011). Exploring the use of social media to measure journal article impact. *AMIA Annual Symposium Proceedings*, (pp.374-381).

Fenner, M. (2014). Altmetrics and Other Novel Measures for Scientific Impact. In S. Bartling & S. Friesike (Eds.), Opening Science (pp. 179–189). Springer International Publishing.

Fienberg, S. E., & Martin, M. E. (1985). Sharing research data. Washington:The National Academies Press.

Franceschini, F., Maisano, D. & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics*, 102 (3), 2181–2186.

Franceschini, F., Maisano, D. & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus. *Journal of Informetrics*, 10 (1), 174–182.

Fuyuno, I., & Cyranoski, D. (2006). Cash for papers: Putting a premium on publication. *Nature*, 441 (7095), 792.

Galligan, F., & Dyas-Correia, Sh. (2013). Altmertics: Rethinking the way we measure. *Serials Review*, 39(1), 56–61.

Galloway, L., & Pease, J. (2013). Altmetrics for the Information Professional: A Primer. *Library and Librarians Publication.* Paper 105. Retrived from http://surface.syr.edu/sul/105

Grimm, J., & Grimm, W. (2004). *The Annotated Brothers Grimm*. (M. Tatar, Ed.) (p. 416). W. W. Norton & Company. Retrieved from http://www.amazon.com/Annotated-Brothers-Grimm-Books/dp/0393058484/ref=pd_sim_b_1

Halfaker, A. & Taraborelli, D. (2015). Wikipedia Scholarly Article Citations. doi:10.6084/m9.figshare.1299540.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 101 (2), 1145-1163.

Haustein, S. (2013). Altmetrics: Present and Future; a Panel discussion at the ASIS&T 2013 Annual Meeting.Montreal, Canada.

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*. 101(2), 1027-1042.

Huvila, I. (2010). Where does the information come from? Information source use patterns in Wikipedia. *Information Research*, 15 (3). http://www.informationr.net/ir/15-3/paper433.html. Accessed 2 May 2017.

Haustein, S. & Sibenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*. 5(3), 446-457.

Haustein, S. (2014). "Readership metrics", Cronin, B. and Sugimoto, Cassidy R. (eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press, Cambridge, pp. 327-344.

Jamali, H. R., Nicholas, D., Watkinson, A., Herman, E., Tenopir, C., Levine, K., Allard, S., Christian, L, Volentine, R., Boehm, R. & Nichols, F. (2014), How scholars implement trust in their reading, citing and publishing activities: Geographical differences. *Library & Information Science Research*, 36 (3-4), 192-202.

Kousha, K., Thelwall, M. & Rezaie, S. (2010). Using the Web for research evaluation: The Integrated Online Impact indicator. *Journal of Informetrics*, 4(1), 124–135.

Kousha, K. & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books?. *Journal of the Association for Information Science and Technology*, 68 (3), 762-779.

Kousha, K. & Thelwall, M. (2017). News stories as ecidence for research? BBC citations from Articles, Books and Wikipedia, *Journal of the Association for Information Science and Technology*, 67 (8), 2017-2028.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter , a Social Network or a News Media? Categories and Subject Descriptors. *Proceedings of the 19th international conference on World wide web*, Raleigh, North Carolina, USA.

Li, X., Thelwall, M., & Giustini, D. (2011), Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91 (2), 461–471.

Lin, J. & Fenner, M. (2014). An analysis of Wikipedia references across PLOS publications. figshare.1048991.v3.

Lindsey, D. (1989), Using citation counts as a measure of quality in science: measuring what's measurable rather than what's valid. *Scientometrics*, 15 (3–4), 189–203.

Liu, J., & Adie, E. (2013). Five challenges in altmetrics: A toolmaker's perspective. *Bulletin of the american society for information science and technology*, 39(4), 31–34.

Luyt, B. & Tan, D. (2010). Improving Wikipedia's credibility: References and citations in a sample of history articles. *Journal of the Association for Information Science and Technology*, 61 (4), 715-722.

Maflahi, N. & Thelwall, M. (2016). When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals. *Journal of the Association for Information Science and Technology*, 67 (1), 191-199.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M. & Delgado López-Cózar, E. (2016). Back to the past: on the shoulders of an academic search engine giant. *Scientometrics*, 107 (3), 1477-1487.

MacRoberts, M. H. & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12.

Meho, L. I., & Yang, K. (2007). Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.

Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. A. & Lanamäki, A. (2015). The sum of all human knowledge: a systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology*, 66 (2), 219-245.

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, 56(10), 1088-1097.

Mohammadi, E., and Thelwall, M. (2014), Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows, *Journal of the Association for Information Science and Technology*, 65(8), 1627–1638.

Mohammadi, E., Thelwall, M., Haustein, S. & Larivière, V. (2015), Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, 66(9), 1832–1846.

Mohammadi, E., Thelwall, M. & Kousha, K. (2016). Can Mendeley bookmarks reflect readership? A survey of user motivations, *Journal of the Association for Information Science and Technology*, 67(5), 1198-1209.

Mongeon, P. & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometric,* 106(1), 213-228.

Nández, G. & Borrego, Á. (2013). Use of social networks for academic purposes: a case study. *Electronic Library*, 31(6), 781–791.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali, H. R. & Polydoratou, P. (2005). Revisiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication. *Information Processing and Management*, 41(6), 1441-1461.

Nielsen, F. (2007). Scientific citations in Wikipedia. *First Monday*, 12(8). http://firstmonday.org/article/view/1997/1872. Accessed 2 May 2017.

Odlyzko, A. M. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15(1), 7-19.

Opthof, T. (1997). Sense and nonsense about the impact factor. *Cardiovascular Research*, 33(1), 1-7.

Pooladian, A., Borrego, Á. (2016). Twenty years of readership of library and information science literature under Mendeley's microscope. *Performance Measurement and Metrics*, 17 (3), 67-77.

Priem, J. (2014). Altmetrics. In C. R. Cronin, Blaise Sugimoto (Ed.), Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact (p. 342). MIT Press.

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the 73rd ASIS&T Annual Meeting*, *73*. Retrieved from http://jasonpriem.org/self-archived/Priem_Costello_Twitter.pdf

Priem, J., & Hemminger, B.H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, 15, 7–5.

Priem, J., Parra, C., Piwowar, H., Groth, P. & Waagmeester, A. (2012). Uncovering impacts: a case study in using altmetrics tools. Workshop on the Semantic *Publishing SePublica 2012 at the 9th Extended Semantic Web Conference*.

Priem, J., Taraborelli, D., Groth, P. & Nylon, C. (2010). Altmetrics: a manifesto. Retrieved 15 April 2014 from http://altmetrics.org/manifesto.

Seglen, PO. (1992). How representative is the journal impact factor?. *Research Evaluation*, 2(3), 143-149.

Seglen, PO. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 497.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K. & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics*, 101(2), 1113-1128.

Stankus, T. & Spiegel, S. E. (2010). Wikipedia, Scholarpedia, and References to Journals in the Brain and Behavioral Sciences: A Comparison of Cited Sources and

Recommended Readings in Matching Free Online Encyclopedia Entries. *Science & Technology Libraries,* 29(3) , 258-265.

Tarango, J., Ascencio-Baca, G., Romo-González, J. R. & Gutiérrez-Balderrama, J. P. (2017). Tendencias de información en la categoría de artículos destacados de Wikipedia: una perspectiva sobre la obsolescencia de los contenidos virtuales y de libre acceso. *Anales de Documentación*, 20(1), 2-19.

Tenopir, C. & King, D. W. (2000). *Towards electronic journals: Realities for scientists, librarians, and publishers*. Washington, DC: Special Libraries Association.

Teixeira, M. C., Thomaz, S. M., Michelan, T. S., Mormul, R. P., Meurer, T., Fasolli, J.V. & Silveira, M. J. (2013). Incorrect Citations Give Unfair Credit to Review Authors in Ecology Journals. *PLoS ONE*, 8(12), e81871.

Teplitskiy, M., Lu, G. & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9), 2116-2127.

Thelwall, M. (2009). *Introduction to webometrics: Quantitative web research for the Social Sciences*. San Rafael, CA: Morgan & Claypool.

Thelwall, M. (2016). Does astronomy research become too dated for the public? Wikipedia citations to astronomy and astrophysics journal articles 1996-2014. *El Profesional de la Información*, 25(6), 893-900.

Thelwall, M. (2016). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? *Journal of Librarianship and Information Science*, 49(2), 144-151.

Thelwall, M. & Sud, P. (2016). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology*, 67(12), 3036-3050.

Thelwall, M. (2017). Does Mendeley provides ecidence of the educational value of journal articles?. *Learned publishing*, 30(2), 107-113.

Thomson Reuters (2013), Web of Science Help: Searching the Document Type Field, available at: http://images.webofknowledge.com/WOKRS59B4/help/WOS/hs_document_type.html (accessed 1 June 2016).

Walters, W. & Wilders, E. (2016). Disciplinary, National, and Departmental Contributions to the Literature of Library and Information Science, 2007–2012, *Journal of the Association for Information Science and Technology*, 67(6), 1487-1506.

Waltman, L., & Costas, R. (2014). F1000 recommendations as a Potentail New Data Source for Research Evaluation: A comparison with citations, *Journal of the American Society for Information Science and Technology*, 65(3), 433-445.

Weale, A. R., Bailey, M., & Lear, P. A. (2004). The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. *BMC Medical Research Methodology*, 4(1), 14.

World Bank (2016). World Development Indicators database: GDP per capita, PPP. http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?order=wbapi_data_value_2014+wbapi_data_value+wbapi_data_value-last&sort=desc. Accessed 14 May 2016.

Wouters, P., & Costas, R. (2012). *Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century*. SURFfoundation. Retrieved February 28, 2013, from www.surf.nl/en/publicaties/pages/users_narcissism_control.aspx

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512 (7513), 126–129.

Vinkler, P. (2007). Eminence of scientists in the light of the h-index and other scientometric indicators. *Journal of Information Science, 33*(4), 481–491.

Zahedi, Z., Costas, R. & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics,* 101(2), 1491-1513.

Zahedi, Z., Costas, R. and Wouters, P. (2017). Mendeley redearship as a filtering tool to identify highly cited publications. *Journal of the association for information science and technology*, 68(10), 2511-2521.

Annexes

Annex 1

Table 14. Differences in readership depending on journal

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
1	*INFORMATION TECHNOLOGY FOR DEVELOPMENT*	111	111	3492	100.00
2	*TELEMATICS AND INFORMATICS*	152	151	3667	99.34
3	*INTERNATIONAL JOURNAL OF COMPUTER-SUPPORTED COLLABORATIVE LEARNING*	167	165	7236	98.80
4	*INFORMATION SYSTEMS JOURNAL*	362	354	15887	97.79
5	*INFORMATION & MANAGEMENT*	1097	1065	47491	97.08
6	*LIBRARY & INFORMATION SCIENCE RESEARCH*	464	450	11378	96.98
7	*INFORMATION SYSTEMS RESEARCH*	595	576	37529	96.81
8	*INFORMATION AND ORGANIZATION*	88	85	3541	96.59
9	*JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*	2041	1958	50458	95.93
10	*INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*	865	829	29305	95.84
11	*INFORMATION PROCESSING & MANAGEMENT*	1288	1230	26049	95.50
12	*JOURNAL OF STRATEGIC INFORMATION SYSTEMS*	290	276	16321	95.17
13	*KNOWLEDGE MANAGEMENT RESEARCH & PRACTICE*	225	213	3353	94.67
14	*ETHICS AND INFORMATION TECHNOLOGY*	151	142	2663	94.04
15	*INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*	1179	1107	25171	93.89
16	*JOURNAL OF INFORMETRICS*	451	422	9032	93.57
17	*SCIENTOMETRICS*	2826	2632	52733	93.14
18	*GOVERNMENT INFORMATION QUARTERLY*	710	661	27571	93.10

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
19	*JOURNAL OF KNOWLEDGE MANAGEMENT*	361	335	16423	92.80
20	*TELECOMMUNICATIONS POLICY*	1098	1006	14273	91.62
21	*INFORMATION DEVELOPMENT*	176	161	2052	91.48
22	*SERIALS REVIEW*	242	221	2970	91.32
23	*JOURNAL OF DOCUMENTATION*	654	596	13040	91.13
24	*INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SYSTEMS*	83	75	1069	90.36
25	*HEALTH INFORMATION AND LIBRARIES JOURNAL*	381	344	6145	90.29
26	*INFORMATION TECHNOLOGY & MANAGEMENT*	163	147	2429	90.18
27	*INFORMATION TECHNOLOGY & PEOPLE*	112	100	2347	89.29
28	*EUROPEAN JOURNAL OF INFORMATION SYSTEMS*	523	466	21372	89.10
29	*ONLINE INFORMATION REVIEW*	643	569	9744	88.49
30	*ASLIB PROCEEDINGS*	687	605	8487	88.06
31	*RESTAURATOR-INTERNATIONAL JOURNAL FOR THE PRESERVATION OF LIBRARY AND ARCHIVAL MATERIAL*	324	284	1060	87.65
32	*LIBRARY COLLECTIONS ACQUISITIONS & TECHNICAL SERVICES*	345	300	2465	86.96
33	*INFORMATION SOCIETY*	436	376	10199	86.24
34	*JOURNAL OF HEALTH COMMUNICATION*	1028	881	14735	85.70
35	*JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*	184	152	1396	82.61
36	*ELECTRONIC LIBRARY*	957	789	11039	82.45
37	*JOURNAL OF ACADEMIC LIBRARIANSHIP*	1151	938	16175	81.49
38	*SOCIAL SCIENCE COMPUTER REVIEW*	722	580	12045	80.33
39	*PROFESIONAL DE LA INFORMACION*	663	531	7801	80.09

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
40	*PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS*	401	321	3742	80.05
41	*JOURNAL OF INFORMATION TECHNOLOGY*	460	365	12805	79.35
42	*SOCIAL SCIENCE INFORMATION SUR LES SCIENCES SOCIALES*	527	416	5350	78.94
43	*LIBRI*	529	409	2688	77.32
44	*JOURNAL OF INFORMATION SCIENCE*	877	675	14098	76.97
45	*JOURNAL OF GLOBAL INFORMATION MANAGEMENT*	146	112	1381	76.71
46	*RESEARCH EVALUATION*	400	306	4208	76.50
47	*REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA*	202	154	1462	76.24
48	*JOURNAL OF COMPUTER-MEDIATED COMMUNICATION*	447	336	27460	75.17
49	*JOURNAL OF ORGANIZATIONAL AND END USER COMPUTING*	83	62	316	74.70
50	*LIBRARY HI TECH*	601	440	7823	73.21
51	*JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE*	370	269	3809	72.70
52	*INTERNATIONAL INFORMATION & LIBRARY REVIEW*	109	75	376	68.81
53	*INTERLENDING & DOCUMENT SUPPLY*	505	331	1532	65.54
54	*PORTAL-LIBRARIES AND THE ACADEMY*	351	229	3771	65.24
55	*LEARNED PUBLISHING*	611	383	3111	62.68
56	*JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION*	651	402	7322	61.75
57	*JOURNAL OF MANAGEMENT INFORMATION SYSTEMS*	585	361	18969	61.71
58	*JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*	2786	1635	48595	58.69
59	*LIBRARY ACQUISITIONS-PRACTICE AND THEORY*	130	76	225	58.46
60	*LIBRARY QUARTERLY*	345	196	2079	56.81
61	*ASIST 2003: PROCEEDINGS OF THE 66TH ASIST ANNUAL MEETING*	119	66	349	55.46

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
62	*LIBRARIES & THE CULTURAL RECORD*	22	12	87	54.55
63	*JOURNAL OF GOVERNMENT INFORMATION*	257	136	1292	52.92
64	*LIBRARY TRENDS*	861	447	5928	51.92
65	*ASLIB JOURNAL OF INFORMATION MANAGEMENT*	34	16	105	47.06
66	*AUSTRALIAN ACADEMIC & RESEARCH LIBRARIES*	121	56	416	46.28
67	*PUBLISHING RESEARCH QUARTERLY*	273	124	572	45.42
68	*PROGRAM-AUTOMATED LIBRARY AND INFORMATION SYSTEMS*	42	19	45	45.24
69	*ASIST 2002: PROCEEDINGS OF THE 65TH ASIST ANNUAL MEETING*	116	52	236	44.83
70	*ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY*	173	77	2616	44.51
71	*IFLA JOURNAL-INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS*	32	14	40	43.75
72	*AUSTRALIAN LIBRARY JOURNAL*	129	53	250	41.09
73	*JOURNAL OF SCHOLARLY PUBLISHING*	391	155	1094	39.64
74	*KNOWLEDGE ORGANIZATION*	354	134	759	37.85
75	*INVESTIGACION BIBLIOTECOLOGICA*	168	62	129	36.90
76	*JOURNAL OF EDUCATION FOR LIBRARY AND INFORMATION SCIENCE*	134	48	140	35.82
77	*LIBRARY RESOURCES & TECHNICAL SERVICES*	396	141	987	35.61
78	*INFORMATION TECHNOLOGY AND LIBRARIES*	448	151	2152	33.71
79	*COLLEGE & RESEARCH LIBRARIES*	673	214	3483	31.80
80	*INFORMATION & CULTURE*	60	19	65	31.67
81	*INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL*	586	185	2478	31.57

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
82	*BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION*	363	107	802	29.48
83	*AFRICAN JOURNAL OF LIBRARY ARCHIVES AND INFORMATION SCIENCE*	111	32	145	28.83
84	*INVESTIGACION BIBLIOTECOLOGIA*	7	2	2	28.57
85	*MIS QUARTERLY*	664	189	20309	28.46
86	*PROCEEDINGS OF THE ASIS ANNUAL MEETING*	154	41	223	26.62
87	*REFERENCE & USER SERVICES QUARTERLY*	515	127	1264	24.66
88	*ONLINE & CDROM REVIEW*	160	39	101	24.38
89	*LIBRARY AND INFORMATION SCIENCE*	131	31	100	23.66
90	*CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REVUE CANADIENNE DES SCIENCES DE L INFORMATION ET DE BIBLIOTHECONOMIE*	199	45	209	22.61
91	*BEHAVIORAL & SOCIAL SCIENCES LIBRARIAN*	54	12	20	22.22
92	*SERIALS LIBRARIAN*	216	48	158	22.22
93	*INTERNATIONAL FORUM ON INFORMATION AND DOCUMENTATION*	65	14	74	21.54
94	*EDUCATION FOR INFORMATION*	49	10	24	20.41
95	*AMERICAN ARCHIVIST*	25	5	19	20.00
96	*PERSPECTIVAS EM CIENCIA DA INFORMACAO*	237	46	296	19.41
97	*LAW LIBRARY JOURNAL*	517	93	539	17.99
98	*JOURNAL OF INFORMATION ETHICS*	142	24	84	16.90
99	*NACHRICHTEN FUR DOKUMENTATION*	71	12	14	16.90
100	*MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE*	168	27	224	16.07
101	*DATA BASE FOR ADVANCES IN INFORMATION SYSTEMS*	131	21	187	16.03
102	*BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*	13	2	2	15.38

Row	Title of Journals	Articles	Articles with Readers	Number of Readers	% Articles with Readers
103	*JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*	566	80	656	14.13
104	*ONLINE*	873	114	635	13.06
105	*INFORMACIOS TARSADALOM*	163	21	50	12.88
106	*JOURNAL OF GLOBAL INFORMATION TECHNOLOGY MANAGEMENT*	57	7	40	12.28
107	*ECONTENT*	753	88	377	11.69
108	*WEB OF KNOWLEDGE - A FESTSCHRIFT IN HONOR OF EUGENE GARFIELD*	27	3	20	11.11
109	*INFORMACAO & SOCIEDADE-ESTUDOS*	218	23	69	10.55
110	*LIBRARY JOURNAL*	2158	217	742	10.06
111	*TRANSINFORMACAO*	135	11	23	8.15
112	*JOURNAL OF THE ASSOCIATION FOR INFORMATION SYSTEMS*	273	19	649	6.96
113	*ASIST 2001: PROCEEDINGS OF THE 64TH ASIST ANNUAL MEETING*	130	9	17	6.92
114	*RQ*	74	5	18	6.76
115	*ZEITSCHRIFT FUR BIBLIOTHEKSWESEN UND BIBLIOGRAPHIE*	445	30	62	6.74
116	*MIS QUARTERLY EXECUTIVE*	112	6	82	5.36
117	*SPECIAL LIBRARIES*	41	2	3	4.88
118	*NFD INFORMATION-WISSENSCHAFT UND PRAXIS*	247	10	15	4.05
119	*SCIENTIST*	3021	98	1140	3.24
120	*DATABASE*	157	5	8	3.18
121	*WILSON LIBRARY BULLETIN*	44	1	1	2.27
122	*CD-ROM PROFESSIONAL*	227	2	2	0.88
123	*DATABASE-THE MAGAZINE OF ELECTRONIC DATABASE REVIEWS*	51	0	0	-
124	*INTERNET WORLD*	391	0	0	-
Total		**54655**	**33295**	**773893**	100.00

Annex 2

List of scientific publications of the thesis

Journals

Pooladian, A., Borrego, Á. (2016). Twenty years of readership of library and information science literature under Mendeley's microscope. *Performance Measurement and Metrics*, 17 (3), pp. 67-77.doi: 10.1108/PMM-02-2016-0006

Pooladian, A., Borrego, Á. (2016). A longitudinal study of the bookmarking of library and information science literature in Mendeley. *Journal of Informetrics*, 10 , pp. 1135-1142.doi : 10.1016/j.joi.2016.10.003

Pooladian, A., Borrego, Á. (2017). Methodological issue in measuring citations in Wikipedia: a case study in Library and Information Science. *Scientometrics*, 17 (3), pp. 455-464.doi: 10.1007/s11192-017-2474-z

The full text of these three publications can be found at the end of this annex.

Presentation in international and national conferences

Pooladian, A., Borrego, Á. (2016). Disseminating knowledge beyond scholarly journals: coverage of Library and Information Science literature in Wikipedia, Bobcatsss2016. Lyon (France), 27-29 January.

Pooladian, A., Borrego, Á. (2015). Readers of Library and Information Science literature by Mendeley, 3rd International Seminar on Libraray and Information Science Education and Research (LIS-ER), Barcelona (Spain), 4-5 June.

Performance Measurement and Metrics

Twenty years of readership of library and information science literature under Mendeley's microscope

Aida Pooladian, Ángel Borrego,

Twenty years of readership of library and information science literature under Mendeley's microscope

Aida Pooladian and Ángel Borrego
Universitat de Barcelona, Barcelona, Spain

Received 4 March 2016
Revised 16 June 2016
Accepted 31 August 2016

Abstract

Purpose – Altmetric indicators have been proposed as a complement to citation counts in research evaluation. Conceivably, they might also be used to supplement other methods for estimating journal readership. The purpose of this paper is to explore the use of Mendeley reference management software to characterize the features of the readership of library and information science (LIS) literature.

Design/methodology/approach – The authors used the Social Sciences Citation Index to retrieve 54,655 articles and reviews published between 1995 and 2014 and indexed in the category "Information Science & Library Science". Each record was then searched in Mendeley to obtain the number of bookmarks of the paper and the academic status of the users.

Findings – Mendeley covers 61 per cent of the LIS literature published in the last 20 years. In all, 75 per cent of the papers published in the last five years had been bookmarked at least once in Mendeley whereas just 55 per cent had been cited in Web of Science. Reviews are bookmarked more frequently than articles, and papers in English have more bookmarks than papers in any other language. Most users of LIS literature are PhD and postgraduate students.

Originality/value – The study shows that altmetrics can be used as a source of data in information behaviour studies. Reference management software provides an unobtrusive means of capturing reading habits in the scholarly literature. Compared to citation counts, bookmarks are rapidly available and also reflect usage outside the academic community.

Keywords Information behaviour, Scholarly communication, Mendeley, Library and information science, Altmetrics, Reference management software

Paper type Research paper

1. Introduction

1.1 Altmetrics as a source of data on academic information behaviour

Altmetrics are non-traditional metrics which have been proposed as an alternative or a complement to traditional citation counts in research assessment. The rationale is that altmetrics cover other aspects of the impact of scientific works, such as the number of views, downloads, bookmarks or mentions in social media. Bornmann (2014) identifies four benefits of altmetrics that help to overcome some of the limitations of citation counts as indicators for research evaluation: their breadth (altmetrics measure impact among non-academic users, such as professionals, whereas citations only measure impact within the academic community); diversity (altmetrics facilitate the evaluation of a wider range of research outcomes beyond publications); speed (altmetrics can measure impact immediately after publication, whereas citations need much more time to accrue), and openness (access to most altmetric data sources is free, whereas commercial citation indexes such as Web of Science or Scopus require a subscription). However, altmetrics also suffer from limitations, including commercialization (resulting in a constant promotion of social media services that can bias altmetric indicators), data quality (including concerns about bias among the user population or lack of standardization), missing evidence (due to the lack of large-scale studies), and the possible manipulation of data.

Readership of academic journals is an informative indicator (though difficult to measure) for the different players involved in scholarly communication. For researchers, the relevance

Performance Measurement and Metrics
Vol. 18 No. 1, 2017
pp. 67-77

1467-8047
DOI 10.1108/PMM-02-2016-0006

of a journal among their intended audience is a key criterion when deciding where to publish (Jamali *et al.*, 2014). Librarians use readership data to make collection management decisions, and editors and publishers use it to monitor the performance of journals in the scholarly marketplace (Haustein, 2014).

In academic and research libraries, methods to measure journals' readership have traditionally included shelving statistics (since journals are usually excluded from loan), document delivery data, participative methods (such as surveys and interviews) and citation counts (Haustein, 2014). The transition from print to electronic journals has increased the amount of data available on the frequency of articles being downloaded. Log files in publishers' servers record journal usage, and initiatives such as the Project COUNTER (projectcounter.org) facilitate the reporting of standardized online usage statistics in a consistent, credible and comparable way across publishers. However, journals' global download usage data are not publicly released by publishers, since this information is commercially sensitive.

Before becoming the standard approach for research evaluation, citation analysis was already used to gather data in studies on information behaviour in academic settings. Analysing the sources cited by scholars in their publications was an unobtrusive way of capturing data on researchers' reading habits that might inform decisions on library collection management. In recent years it has been suggested that altmetrics (i.e. web-based metrics such as views, downloads or mentions in social media) might complement citation analysis in research evaluation procedures. We propose that, in a similar fashion, altmetric indicators might also be used in library settings to explore the information needs and behaviour of library users. The rationale is that the number of users who bookmark an article in an online reference management software is a potential indicator of the readership of the article. This approach has already been explored, with promising results (Haustein and Siebenlist, 2011; Borrego and Fry, 2012).

The use of altmetric data to estimate journal readership might supplement citation analysis in much the same way as a complement to citation counts in research evaluation. Two of the main limitations of citation analysis when estimating journal readership also apply to research assessment: that is, citations only capture readership among authors who publish and therefore cite, and they also take a long time to accumulate. Altmetric data can be helpful for estimating readership outside the academic community; another advantage is that they are available shortly after publication, since papers can attract readers before they are cited (Bornmann and Haunschild, 2015; Maflahi and Thelwall, 2016; Thelwall and Sud, 2016).

1.2 Reference management software

Today, researchers have moved a great deal of their research activity to the web where academic social networking sites allow them to disseminate, comment and collaborate with colleagues (Van Noorden, 2014). Specifically, reference management software allows scholars to record and share their bibliographic references. Mendeley is an example of a service of this kind that help scholars manage, tag, cite and share academic papers. Researchers may also use the tool to create a public profile in order to disseminate their publication output and establish links with scholars in their field. Mendeley counts how frequently an article is "read" (i.e. how often it is bookmarked by Mendeley users), thus providing information on how academics interact with scholarly information. Thus, the number of Mendeley users who bookmark a given article or journal potentially indicates its readership size (Li *et al.*, 2011).

Although Mendeley employs the term "readers" to refer to counts of users of a given paper, throughout this article we use the terms "users" and "bookmarks" to refer to the same concept. The terms "usage" and "bookmarks" describe more precisely the behaviour of the users, since saving an article does not necessarily mean that the user will read it. This is a

feature that, to a certain extent, altmetric data share with citation counts, since citing an article does not necessarily mean that it has been read. Nevertheless it seems likely that more frequently saved articles are more frequently read (Li *et al.*, 2011) and, therefore, bookmark counts may be indicative of the readership of an article or journal. At the moment there is a gap in the literature, since very few studies have addressed the issue of whether Mendeley users bookmark publications in order to read them or not. The only empirical evidence on this issue is a recent survey of 860 Mendeley users (Mohammadi *et al.*, 2016) which showed that 27 per cent had read or intended to read all of their bookmarked records and that 55 per cent had read or intended to read at least half.

It is likely that only a share of the readers of an article will save it in an online reference manager and, more specifically, in Mendeley. However, compared to other reference management software, Mendeley records more users and more bookmarks. In 2013, Mendeley announced that it had reached 2.5 million users (Bonasio, 2013). According to Li *et al.* (2011), around 60 per cent of a sample of *Nature* and *Science* articles were stored in CiteULike whereas more than 90 per cent were in Mendeley. Further research also indicates that Mendeley provides wider coverage of scholarly journal articles than any other altmetric data source. In all, 63 per cent of the articles indexed in Web of Science between 2005 and 2011 were included in a Mendeley account by April 2013 (Zahedi *et al.*, 2014). However, there are clear differences between the disciplines. Thus, Mohammadi and Thelwall (2014) situated the coverage of Mendeley at 58 per cent of 62,600 articles in the Social Sciences indexed by Web of Science in 2008, whereas the coverage of 14,600 articles in Humanities decreased to 28 per cent. Another study showed a coverage of 72 per cent for a sample of articles published in clinical medicine in 2008, but this figure was considerably lower for articles in the social sciences (47 per cent), engineering and technology (35 per cent), chemistry (34 per cent) and physics (31 per cent) (Mohammadi *et al.*, 2015).

In addition to the number of bookmarks, Mendeley also provides an estimate of who the users of the articles are. Mendeley libraries are anonymous, making it impossible to know which papers (or even how many papers) have been bookmarked by a given user. However, Mendeley provides collective demographic details about the users of a given paper – discipline, country of residence and academic status – based on the information these users provide when creating their accounts. Previous research indicates that most Mendeley users are PhD students, postgraduates and postdoctoral researchers (Mohammadi *et al.*, 2015). However, it should be noted that demographic status is self-reported and may therefore be inaccurate.

1.3 Objectives

This paper aims to explore the use of reference management software to estimate the readership of academic literature. Specifically, it uses Mendeley data to explore the readership of library and information science (LIS) literature published in the last 20 years, from 1995 to 2014. This analysis is of interest because bookmarking data can illustrate a different role compared to citations in order to estimate the readership of scholarly literature. The research is underpinned by the following questions:

RQ1. What was Mendeley's coverage in 2015 of the LIS literature published in the 20-year period 1995-2014?

RQ2. What was the share of the LIS literature bookmarked in Mendeley compared to the share of the same literature cited in Web of Science?

RQ3. Did the number of bookmarks in Mendeley vary according to the type of paper (article or review)?

RQ4. Did the number of bookmarks in Mendeley vary according to the language of the paper?

RQ5. What is the academic status of Mendeley users of LIS literature?

RQ6. What are the most widely bookmarked journals in LIS according to Mendeley figures?

2. Methods

In the first quarter of 2015 we used the Social Sciences Citation Index to retrieve a total of 54,778 papers published between 1995 and 2014 and indexed in the category "Information Science & Library Sciences". The search was limited to articles and reviews, and our analysis excluded all other types of document, such as book reviews, editorials, letters or proceedings. Web of Science defines articles as "reports of research on original works" whereas a review is "a renewed study of material previously studied" (Thomson Reuters, 2013). Throughout this article we use the term "papers" to refer to the set of both articles and reviews.

When this initial search was completed, between 27 April and 8 May 2015 each retrieved record was searched in Mendeley using Webometric Analyst software (Thelwall, 2009) with a query containing the title of the paper, the first author's last name, the year of publication and the digital object identifier.

The software was unable to search 123 records, probably because they were incomplete or contained errors generated while being downloaded from Web of Science. These records were subsequently removed from the analysis, leaving a final sample of 54,655 papers. For each paper, the journal, language and year of publication were obtained from Web of Science and the number of users and their academic status were recorded from Mendeley. (Note that Mendeley only returns the three main academic status categories for the users of each paper).

3. Results

3.1 Mendeley coverage of LIS literature

Of the 54,655 LIS papers published between 1995 and 2014, 33,295 (61 per cent) had been bookmarked in Mendeley by at least one user by May 2015 (Table I). Since the first public beta version of Mendeley was released in August 2008, it was expected that the coverage of the literature would be related to its publication year. Thus, as shown in Figure 1, nearly two-thirds (64 per cent) of the "older" papers – those published between 1995 and 1999 – had no bookmarks. However, the number of articles and reviews with and without bookmarks began to balance in documents published between 2000 and 2004, and the number of papers bookmarked then continued to increase, so that more than 70 per cent of the papers published in the last decade (2005-2014) were bookmarked at least once in Mendeley by May 2015.

One of the possible advantages of bookmark counts for estimating readership is that they may be able to capture usage immediately after publication, whereas citations need much more time to accrue. Figure 2 compares the share of LIS literature bookmarked in Mendeley with the share of papers cited in Web of Science by year of publication. Whereas the percentage of cited papers is higher among papers published between 1995 and 2004,

Table I. Mendeley coverage of LIS literature published between 1995 and 2014

	Papers	% of papers
Bookmarked	33,295	60.9
Without bookmarks	21,360	39.1
Total	54,655	100.0

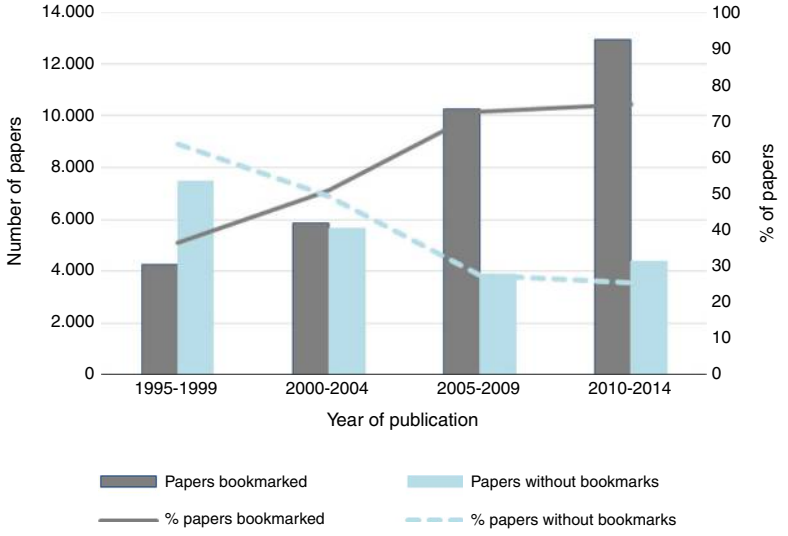

Figure 1.
Mendeley coverage of LIS literature by year of publication

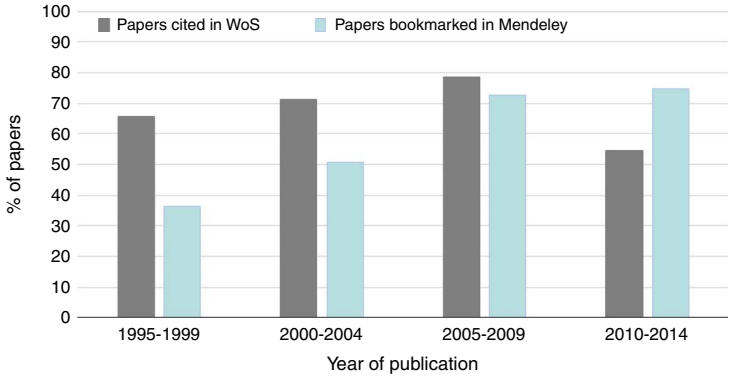

Figure 2.
LIS literature cited and bookmarked by year of publication

the share of cited and bookmarked papers becomes progressively balanced for papers published between 2005 and 2009. In the case of papers published in the last five years (2010-2014), 55 per cent had been cited in Web of Science at the time of data collection, in May 2015, whereas 75 per cent had been bookmarked at least once in Mendeley.

For each paper, Mendeley provides the number of "readers", i.e., the number of users who bookmarked the paper in their libraries. As shown in Table II, one-quarter of the papers (26 per cent) had between one and five users and over half (56 per cent) had between one and 15. At the other end of the scale, an article entitled "Social network sites: definition, history, and scholarship" published in 2007 had 10,217 users. However, this case was particularly extreme, since the second article in the ranking had 893 users.

The distributions of both bookmarks and citations among papers were skewed, since the top 20 per cent of the papers accounted for 75 per cent of the bookmarks and 82 per cent of the citations (Figure 3). However, the top papers by number of bookmarks and by number of citations are not necessarily the same. If we compare the core articles by number of bookmarks (i.e. those with 21 or more bookmarks each) and the core

	Papers	% of papers
1-5 users	8,537	25.6
6-10 users	5,742	17.2
11-15 users	4,380	13.2
16-20 users	3,198	9.6
21-25 users	2,389	7.2
26-30 users	1,701	5.1
31-35 users	1,338	4.0
36-40 users	1,021	3.1
41-45 users	859	2.6
46-50 users	688	2.1
>50 users	3,442	10.3
Total	33,295	100.0

Table II.
Number of users per paper

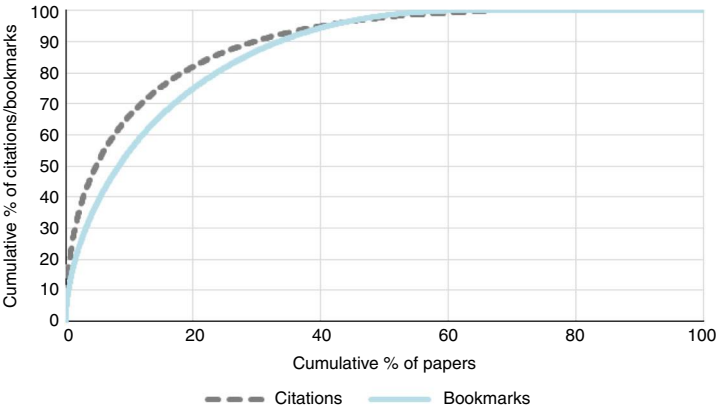

Figure 3.
Distributions of citations and bookmarks

articles by number of citations (i.e. those with ten or more citations each), we observe an overlap of 53 per cent. This means that nearly half of the articles among those with the highest number of bookmarks were not in the top by number citations and vice versa.

3.2 Readership by type of paper (article or review)

In total, 67 per cent of the LIS reviews published between 1995 and 2014 were bookmarked in Mendeley, against 61 per cent of the articles. The median and the average number of bookmarks per review were 2.5 times higher than the number of bookmarks per article (Table III). These differences were similar to those observed regarding citations, with a larger share of reviews being cited (81 per cent) compared to articles (66 per cent). In this case the median and the average number of citations per review were four times the median and the average number of citations per article.

3.3 Readership by language

The number of bookmarks varied according to the language of the paper. As shown in Table IV, the overwhelming majority (95 per cent) of the LIS papers indexed in the Social Sciences Citation Index between 1995 and 2014 were in English. The second language by number of papers was Spanish. Overall, the percentages of English and Spanish papers bookmarked in Mendeley were similar. The number of LIS papers in other languages

	Articles ($n = 53,276$)	Reviews ($n = 1,379$)
Bookmarks		
Papers bookmarked	32,370 (60.8%)	925 (67.1%)
Q1 (bookmarks)	0	0
Median (bookmarks)	4	10
Q3 (bookmarks)	17	39
Average (bookmarks)	13.6	34.0
SD (bookmarks)	34.0	67.0
Citations		
Papers cited	35,349 (66.4%)	1,116 (80.9%)
Q1 (citations)	0	1
Median (citations)	2	8
Q3 (citations)	7	27
Average (citations)	8.2	33.6
SD (citations)	28.6	98.6

Table III. Bookmarks and citations by type of paper (article or review)

Language	Papers	Papers bookmarked	% of papers bookmarked	Average number of bookmarks (SD)	Median number of bookmarks
English	51,912	32,360	62.3	14.7 (53.7)	4
Spanish	1,041	720	69.2	8.6 (15.7)	4
German	716	42	5.9	0.1 (0.4)	0
Portuguese	548	68	12.4	0.6 (3.5)	0
Hungarian	162	21	13.0	0.3 (1.3)	0
French	150	54	36.0	1.2 (2.4)	0
Japanese	118	27	22.9	0.7 (1.7)	0
Other	8	3	37.5	4.1 (9.7)	0
Total	54,655	33,295	60.9	14.2 (52.5)	4

Table IV. Readership by language

indexed in the Social Sciences Citation Index was much smaller and, in all cases, the percentage of papers bookmarked in Mendeley and the average and median number of bookmarks per paper were smaller than among English and Spanish papers.

3.4 Academic status of the users of LIS literature

The largest group of users of LIS literature in Mendeley were PhD students (34 per cent), followed by postgraduate students (22 per cent) and librarians (8 per cent), as shown in Table V.

3.5 Readership by journal

The papers in the sample had been published in 124 journals. Nearly all of them had articles bookmarked, although there were notable differences in the percentage of papers with and without bookmarks (Table VI). In the case of one-fifth (19 per cent) of the journals, nearly all the papers had been bookmarked at least once. However, in the case of one-quarter of other journals, a much smaller share of the papers (less than 20 per cent) had been bookmarked. Table VII details the top 23 journals by percentage of papers bookmarked.

4. Discussion

Studies of readership of academic journals are of interest to the different stakeholders involved in scholarly communication: authors, editors, publishers, librarians, etc. Methods to measure journal readership have traditionally ranged from participative methods such as

Occupation provided by the Mendeley API	Merged	%
PhD student	PhD student	33.8
Doctoral student		
Student (postgraduate)	Postgraduate student	22.2
Student (master)		
Librarian	Librarian	7.5
Assistant professor	Assistant professor	6.9
Lecturer		
Researcher (at an academic institution)	Researcher (at an academic institution)	5.7
Associate professor	Associate professor	5.3
Senior lecturer		
Student (bachelor)	Student (bachelor)	5.0
Professor	Professor	4.2
Postdoc	Postdoc	4.2
Other professional	Other professional	3.3
Researcher (at a non-academic institution)	Researcher (at a non-academic institution)	2.0

Table V.
Users by academic status

% of papers with users	Number of journals	% of journals
91-100	23	18.5
81-90	14	11.3
71-80	14	11.3
61-70	6	4.8
51-60	7	5.6
41-50	8	6.5
31-40	9	7.3
21-30	12	9.7
11-20	15	12.1
1-10	14	11.3
0	2	1.6
Total	124	100

Table VI.
Readership by journal

surveys and interviews, to unobtrusive methods such as shelving statistics, document delivery data, download figures and citation counts. Recently, altmetric indicators have been proposed as a complement to citation metrics for research evaluation since they are available sooner and can capture impact beyond the academic community. In a similar fashion, altmetric indicators such as bookmarks in reference management software might be used to supplement other methods for estimating journal readership.

The results of this study confirm the viability of using bookmark counts to explore the behaviour of users of scholarly literature. The coverage of LIS literature by Mendeley is very extensive. By the first quarter of 2015, Mendeley covered 61 per cent of the LIS literature indexed in the Social Sciences Citation Index in the previous 20 years. This is higher than the percentage observed in previous research, which situated Mendeley's coverage of the social sciences literature somewhere between 47 and 58 per cent (Mohammadi and Thelwall, 2014; Mohammadi *et al.*, 2015). Given that our research analysed a longer period of time than previous studies focussing on more recent literature, the higher coverage observed in this study is particularly significant. Actually, the Mendeley coverage for LIS papers published in the last decade (2005-2014) was higher than 70 per cent. In a similar fashion, although the percentage of LIS articles cited in Web of Science is higher than the percentage of LIS articles bookmarked in Mendeley for

Journal	Papers published	Papers bookmarked	% of papers bookmarked	JCR quartile 2014
Information Technology for Development	111	111	100.0	Q3
Telematics and Informatics	152	151	99.3	Q2
International Journal of Computer-Supported Collaborative Learning	167	165	98.8	Q1
Information Systems Journal	362	354	97.8	Q1
Information & Management	1,097	1,065	97.1	Q1
Library & Information Science Research	464	450	97.0	Q2
Information Systems Research	595	576	96.8	Q1
Information and Organization	88	85	96.6	Q1
Journal of the American Society for Information Science and Technology	2,041	1,958	95.9	Q1
International Journal of Information Management	865	829	95.8	Q1
Information Processing & Management	1,288	1,230	95.5	Q2
Journal of Strategic Information Systems	290	276	95.2	Q1
Knowledge Management Research & Practice	225	213	94.7	Q3
Ethics and Information Technology	151	142	94.0	Q2
International Journal of Geographical Information Science	1,179	1,107	93.9	Q1
Journal of Informetrics	451	422	93.6	Q1
Scientometrics	2,826	2,632	93.1	Q1
Government Information Quarterly	710	661	93.1	Q1
Journal of Knowledge Management	361	335	92.8	Q1
Telecommunications Policy	1,098	1,006	91.6	Q2
Information Development	176	161	91.5	Q3
Serials Review	242	221	91.3	Q3
Journal of Documentation	654	596	91.1	Q2

Table VII.
Top journals by percentage of papers bookmarked

papers published before 2009, the situation is reversed for papers published since 2010. This result confirms that altmetrics offers the important advantage of speed compared to citations for estimating readership.

In addition to speed, another possible advantage of altmetric indicators over citation counts is that they can help to estimate impact outside the academic community. This is especially interesting in fields such as LIS where practicing professionals account for nearly a quarter of the literature published worldwide (Walters and Wilders, 2016). Our results show that, for LIS scholarly literature, librarians and other professionals account for more than 10 per cent of the bookmarks, even though they still come behind PhD and postgraduate students. The number of librarians who use the tool to manage professional literature may also help to explain the higher coverage of LIS literature in Mendeley compared to other disciplines in other social sciences, since in other disciplines most readers are academics – mainly students but also faculty and researchers (Mohammadi *et al.*, 2015).

The distribution of both bookmarks and citations among papers is skewed, but the rankings of papers by number of bookmarks and by number of citations are only partly coincidental. That means that there are articles which have a high number of bookmarks but are not among the top papers by number of citations and vice versa. The reasons for these discrepancies may include (in addition to the different life-cycles of bookmarks and citations) the presence of articles that attract communities which use research but do not cite it, such as students and professionals (Thelwall, 2016).

The results also show that the share of reviews bookmarked in Mendeley is higher than that of articles. Additionally, the average and median number of readers per document is higher for reviews than for articles. This is consistent with previous studies that show that

literature reviews are more often cited than regular articles, most likely due to their breadth (Teixeira *et al.*, 2013).

English is the lingua franca of science, and LIS is no exception to this rule. More than 95 per cent of the LIS literature indexed in the Social Sciences Citation Index is in English. In terms of readership, papers in English are more often bookmarked in Mendeley than papers in any other language. Additionally, there are clear differences in the bookmarking of LIS journals. The share of bookmarked articles varies dramatically from journal to journal: some journals have no articles bookmarked at all, while others have nearly all articles bookmarked by at least one reader.

5. Conclusions

The results of this study show that altmetrics can be used as a source of data in information behaviour studies. Reference management software provides an unobtrusive means of capturing reading habits in scholarly literature that are useful to all the stakeholders in the scholarly communication system.

The application of altmetric indicators to supplement citations counts in order to estimate readership presents two advantages over the use of citations alone. Bookmarks are available sooner, as shown by the fact that the percentage of recent literature bookmarked in Mendeley is much higher than the share of literature cited. Additionally, bookmarks are useful to capture usage beyond the academic community, since reference management software can be employed by professionals to manage the literature. This is especially relevant in fields such as LIS where practicing professionals account for a large part of the use of the literature.

References

Bonasio, A. (2013), "Mendeley has 2.5 million users!", *Mendeley Blog*, available at: https://blog.mendeley.com/2013/09/03/mendeley-has-2-5-million-users/ (accessed 1 June 2016).

Bornmann, L. (2014), "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics", *Journal of Informetrics*, Vol. 8 No. 4, pp. 895-903.

Bornmann, L. and Haunschild, R. (2015), "Which people use which scientific papers? An evaluation of data from F1000 and Mendeley", *Journal of Informetrics*, Vol. 9 No. 3, pp. 477-487.

Borrego, A. and Fry, J. (2012), "Measuring researchers' use of scholarly information through social bookmarking data: a case study of BibSonomy", *Journal of Information Science*, Vol. 38 No. 3, pp. 297-308.

Haustein, S. (2014), "Readership metrics", in Cronin, B. and Sugimoto, C.R. (Eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press, Cambridge, pp. 327-344.

Haustein, S. and Siebenlist, T. (2011), "Applying social bookmarking data to evaluate journal usage", *Journal of Informetrics*, Vol. 5 No. 3, pp. 446-457, available at: http://dx.doi.org/10.1016/j.joi.2011.04.002

Jamali, H.R., Nicholas, D., Watkinson, A., Herman, E., Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R. and Nichols, F. (2014), "How scholars implement trust in their reading, citing and publishing activities: geographical differences", *Library & Information Science Research*, Vol. 36 Nos 3-4, pp. 192-202.

Li, X., Thelwall, M. and Giustini, D. (2011), "Validating online reference managers for scholarly impact measurement", *Scientometrics*, Vol. 91 No. 2, pp. 461-471.

Maflahi, N. and Thelwall, M. (2016), "When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 1, pp. 191-199.

Mohammadi, E. and Thelwall, M. (2014), "Mendeley readership altmetrics for the social sciences and humanities: research evaluation and knowledge flows", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 8, pp. 1627-1638.

Mohammadi, E., Thelwall, M. and Kousha, K. (2016), "Can mendeley bookmarks reflect readership? A survey of user motivations", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 5, pp. 1198-1209.

Mohammadi, E., Thelwall, M., Haustein, S. and Larivière, V. (2015), "Who reads research articles? An altmetrics analysis of Mendeley user categories", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 9, pp. 1832-1846.

Teixeira, M.C., Thomaz, S.M., Michelan, T.S., Mormul, R.P., Meurer, T., Fasolli, J.V. and Silveira, M.J. (2013), "Incorrect citations give unfair credit to review authors in ecology journals", *PLoS ONE*, Vol. 8 No. 12, p. e81871, available at: http://dx.doi.org/10.1371/journal.pone.0081871

Thelwall, M. (2009), *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*, Morgan & Claypool, San Rafael, CA.

Thelwall, M. (2016), "Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa?", *Journal of Librarianship and Information Science*, available at: http://doi.org/10.1177/0961000615594867

Thelwall, M. and Sud, P. (2016), "Mendeley readership counts: an investigation of temporal and disciplinary differences", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 12, pp. 3036-3050.

Thomson Reuters (2013), "Web of science help: searching the document type field", available at: http://images.webofknowledge.com/WOKRS59B4/help/WOS/hs_document_type.html (accessed 1 June 2016).

Van Noorden, R. (2014), "Online collaboration: scientists and the social network", *Nature*, Vol. 512 No. 7513, pp. 126-129.

Walters, W. and Wilders, E. (2016), "Disciplinary, national, and departmental contributions to the literature of library and information science, 2007-2012", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 6, pp. 1487-1506.

Zahedi, Z., Costas, R. and Wouters, P. (2014), "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.

About the authors

Aida Pooladian is a PhD Student in the Faculty of Library and Information Science at the University of Barcelona (Spain). Her research interests are altmetrics, scientometrics, social media, social networks and webometrics.

Ángel Borrego is a Lecturer in the Department of Library and Information Science at the University of Barcelona (Spain). He teaches courses in Scientific Information and Research Methods. His research interests focus on scholarly communication and scientists' information behaviour. Ángel Borrego is the corresponding author and can be contacted at: borrego@ub.edu

Performance Measurement and Metrics

Twenty years of readership of library and information science literature under Mendeley's microscope

Aida Pooladian, Ángel Borrego,

Twenty years of readership of library and information science literature under Mendeley's microscope

Aida Pooladian and Ángel Borrego
Universitat de Barcelona, Barcelona, Spain

Received 4 March 2016
Revised 16 June 2016
Accepted 31 August 2016

Abstract

Purpose – Altmetric indicators have been proposed as a complement to citation counts in research evaluation. Conceivably, they might also be used to supplement other methods for estimating journal readership. The purpose of this paper is to explore the use of Mendeley reference management software to characterize the features of the readership of library and information science (LIS) literature.

Design/methodology/approach – The authors used the Social Sciences Citation Index to retrieve 54,655 articles and reviews published between 1995 and 2014 and indexed in the category "Information Science & Library Science". Each record was then searched in Mendeley to obtain the number of bookmarks of the paper and the academic status of the users.

Findings – Mendeley covers 61 per cent of the LIS literature published in the last 20 years. In all, 75 per cent of the papers published in the last five years had been bookmarked at least once in Mendeley whereas just 55 per cent had been cited in Web of Science. Reviews are bookmarked more frequently than articles, and papers in English have more bookmarks than papers in any other language. Most users of LIS literature are PhD and postgraduate students.

Originality/value – The study shows that altmetrics can be used as a source of data in information behaviour studies. Reference management software provides an unobtrusive means of capturing reading habits in the scholarly literature. Compared to citation counts, bookmarks are rapidly available and also reflect usage outside the academic community.

Keywords Information behaviour, Scholarly communication, Mendeley, Library and information science, Altmetrics, Reference management software

Paper type Research paper

1. Introduction

1.1 Altmetrics as a source of data on academic information behaviour

Altmetrics are non-traditional metrics which have been proposed as an alternative or a complement to traditional citation counts in research assessment. The rationale is that altmetrics cover other aspects of the impact of scientific works, such as the number of views, downloads, bookmarks or mentions in social media. Bornmann (2014) identifies four benefits of altmetrics that help to overcome some of the limitations of citation counts as indicators for research evaluation: their breadth (altmetrics measure impact among non-academic users, such as professionals, whereas citations only measure impact within the academic community); diversity (altmetrics facilitate the evaluation of a wider range of research outcomes beyond publications); speed (altmetrics can measure impact immediately after publication, whereas citations need much more time to accrue), and openness (access to most altmetric data sources is free, whereas commercial citation indexes such as Web of Science or Scopus require a subscription). However, altmetrics also suffer from limitations, including commercialization (resulting in a constant promotion of social media services that can bias altmetric indicators), data quality (including concerns about bias among the user population or lack of standardization), missing evidence (due to the lack of large-scale studies), and the possible manipulation of data.

Readership of academic journals is an informative indicator (though difficult to measure) for the different players involved in scholarly communication. For researchers, the relevance

Performance Measurement and Metrics
Vol. 18 No. 1, 2017
pp. 67-77

1467-8047
DOI 10.1108/PMM-02-2016-0006

of a journal among their intended audience is a key criterion when deciding where to publish (Jamali *et al.*, 2014). Librarians use readership data to make collection management decisions, and editors and publishers use it to monitor the performance of journals in the scholarly marketplace (Haustein, 2014).

In academic and research libraries, methods to measure journals' readership have traditionally included shelving statistics (since journals are usually excluded from loan), document delivery data, participative methods (such as surveys and interviews) and citation counts (Haustein, 2014). The transition from print to electronic journals has increased the amount of data available on the frequency of articles being downloaded. Log files in publishers' servers record journal usage, and initiatives such as the Project COUNTER (projectcounter.org) facilitate the reporting of standardized online usage statistics in a consistent, credible and comparable way across publishers. However, journals' global download usage data are not publicly released by publishers, since this information is commercially sensitive.

Before becoming the standard approach for research evaluation, citation analysis was already used to gather data in studies on information behaviour in academic settings. Analysing the sources cited by scholars in their publications was an unobtrusive way of capturing data on researchers' reading habits that might inform decisions on library collection management. In recent years it has been suggested that altmetrics (i.e. web-based metrics such as views, downloads or mentions in social media) might complement citation analysis in research evaluation procedures. We propose that, in a similar fashion, altmetric indicators might also be used in library settings to explore the information needs and behaviour of library users. The rationale is that the number of users who bookmark an article in an online reference management software is a potential indicator of the readership of the article. This approach has already been explored, with promising results (Haustein and Siebenlist, 2011; Borrego and Fry, 2012).

The use of altmetric data to estimate journal readership might supplement citation analysis in much the same way as a complement to citation counts in research evaluation. Two of the main limitations of citation analysis when estimating journal readership also apply to research assessment: that is, citations only capture readership among authors who publish and therefore cite, and they also take a long time to accumulate. Altmetric data can be helpful for estimating readership outside the academic community; another advantage is that they are available shortly after publication, since papers can attract readers before they are cited (Bornmann and Haunschild, 2015; Maflahi and Thelwall, 2016; Thelwall and Sud, 2016).

1.2 Reference management software

Today, researchers have moved a great deal of their research activity to the web where academic social networking sites allow them to disseminate, comment and collaborate with colleagues (Van Noorden, 2014). Specifically, reference management software allows scholars to record and share their bibliographic references. Mendeley is an example of a service of this kind that help scholars manage, tag, cite and share academic papers. Researchers may also use the tool to create a public profile in order to disseminate their publication output and establish links with scholars in their field. Mendeley counts how frequently an article is "read" (i.e. how often it is bookmarked by Mendeley users), thus providing information on how academics interact with scholarly information. Thus, the number of Mendeley users who bookmark a given article or journal potentially indicates its readership size (Li *et al.*, 2011).

Although Mendeley employs the term "readers" to refer to counts of users of a given paper, throughout this article we use the terms "users" and "bookmarks" to refer to the same concept. The terms "usage" and "bookmarks" describe more precisely the behaviour of the users, since saving an article does not necessarily mean that the user will read it. This is a

feature that, to a certain extent, altmetric data share with citation counts, since citing an article does not necessarily mean that it has been read. Nevertheless it seems likely that more frequently saved articles are more frequently read (Li *et al.*, 2011) and, therefore, bookmark counts may be indicative of the readership of an article or journal. At the moment there is a gap in the literature, since very few studies have addressed the issue of whether Mendeley users bookmark publications in order to read them or not. The only empirical evidence on this issue is a recent survey of 860 Mendeley users (Mohammadi *et al.*, 2016) which showed that 27 per cent had read or intended to read all of their bookmarked records and that 55 per cent had read or intended to read at least half.

It is likely that only a share of the readers of an article will save it in an online reference manager and, more specifically, in Mendeley. However, compared to other reference management software, Mendeley records more users and more bookmarks. In 2013, Mendeley announced that it had reached 2.5 million users (Bonasio, 2013). According to Li *et al.* (2011), around 60 per cent of a sample of *Nature* and *Science* articles were stored in CiteULike whereas more than 90 per cent were in Mendeley. Further research also indicates that Mendeley provides wider coverage of scholarly journal articles than any other altmetric data source. In all, 63 per cent of the articles indexed in Web of Science between 2005 and 2011 were included in a Mendeley account by April 2013 (Zahedi *et al.*, 2014). However, there are clear differences between the disciplines. Thus, Mohammadi and Thelwall (2014) situated the coverage of Mendeley at 58 per cent of 62,600 articles in the Social Sciences indexed by Web of Science in 2008, whereas the coverage of 14,600 articles in Humanities decreased to 28 per cent. Another study showed a coverage of 72 per cent for a sample of articles published in clinical medicine in 2008, but this figure was considerably lower for articles in the social sciences (47 per cent), engineering and technology (35 per cent), chemistry (34 per cent) and physics (31 per cent) (Mohammadi *et al.*, 2015).

In addition to the number of bookmarks, Mendeley also provides an estimate of who the users of the articles are. Mendeley libraries are anonymous, making it impossible to know which papers (or even how many papers) have been bookmarked by a given user. However, Mendeley provides collective demographic details about the users of a given paper – discipline, country of residence and academic status – based on the information these users provide when creating their accounts. Previous research indicates that most Mendeley users are PhD students, postgraduates and postdoctoral researchers (Mohammadi *et al.*, 2015). However, it should be noted that demographic status is self-reported and may therefore be inaccurate.

1.3 Objectives

This paper aims to explore the use of reference management software to estimate the readership of academic literature. Specifically, it uses Mendeley data to explore the readership of library and information science (LIS) literature published in the last 20 years, from 1995 to 2014. This analysis is of interest because bookmarking data can illustrate a different role compared to citations in order to estimate the readership of scholarly literature. The research is underpinned by the following questions:

RQ1. What was Mendeley's coverage in 2015 of the LIS literature published in the 20-year period 1995-2014?

RQ2. What was the share of the LIS literature bookmarked in Mendeley compared to the share of the same literature cited in Web of Science?

RQ3. Did the number of bookmarks in Mendeley vary according to the type of paper (article or review)?

RQ4. Did the number of bookmarks in Mendeley vary according to the language of the paper?

RQ5. What is the academic status of Mendeley users of LIS literature?

RQ6. What are the most widely bookmarked journals in LIS according to Mendeley figures?

2. Methods

In the first quarter of 2015 we used the Social Sciences Citation Index to retrieve a total of 54,778 papers published between 1995 and 2014 and indexed in the category "Information Science & Library Sciences". The search was limited to articles and reviews, and our analysis excluded all other types of document, such as book reviews, editorials, letters or proceedings. Web of Science defines articles as "reports of research on original works" whereas a review is "a renewed study of material previously studied" (Thomson Reuters, 2013). Throughout this article we use the term "papers" to refer to the set of both articles and reviews.

When this initial search was completed, between 27 April and 8 May 2015 each retrieved record was searched in Mendeley using Webometric Analyst software (Thelwall, 2009) with a query containing the title of the paper, the first author's last name, the year of publication and the digital object identifier.

The software was unable to search 123 records, probably because they were incomplete or contained errors generated while being downloaded from Web of Science. These records were subsequently removed from the analysis, leaving a final sample of 54,655 papers. For each paper, the journal, language and year of publication were obtained from Web of Science and the number of users and their academic status were recorded from Mendeley. (Note that Mendeley only returns the three main academic status categories for the users of each paper).

3. Results

3.1 Mendeley coverage of LIS literature

Of the 54,655 LIS papers published between 1995 and 2014, 33,295 (61 per cent) had been bookmarked in Mendeley by at least one user by May 2015 (Table I). Since the first public beta version of Mendeley was released in August 2008, it was expected that the coverage of the literature would be related to its publication year. Thus, as shown in Figure 1, nearly two-thirds (64 per cent) of the "older" papers – those published between 1995 and 1999 – had no bookmarks. However, the number of articles and reviews with and without bookmarks began to balance in documents published between 2000 and 2004, and the number of papers bookmarked then continued to increase, so that more than 70 per cent of the papers published in the last decade (2005-2014) were bookmarked at least once in Mendeley by May 2015.

One of the possible advantages of bookmark counts for estimating readership is that they may be able to capture usage immediately after publication, whereas citations need much more time to accrue. Figure 2 compares the share of LIS literature bookmarked in Mendeley with the share of papers cited in Web of Science by year of publication. Whereas the percentage of cited papers is higher among papers published between 1995 and 2004,

Table I. Mendeley coverage of LIS literature published between 1995 and 2014

	Papers	% of papers
Bookmarked	33,295	60.9
Without bookmarks	21,360	39.1
Total	54,655	100.0

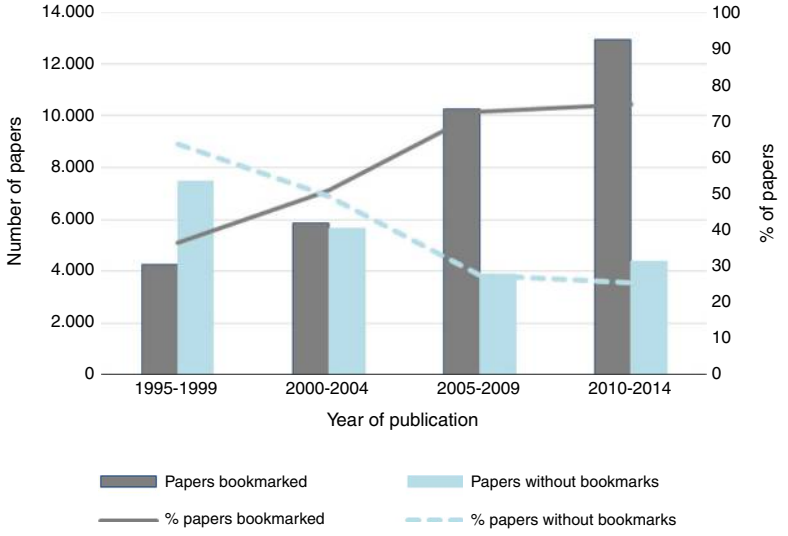

Figure 1.
Mendeley coverage of LIS literature by year of publication

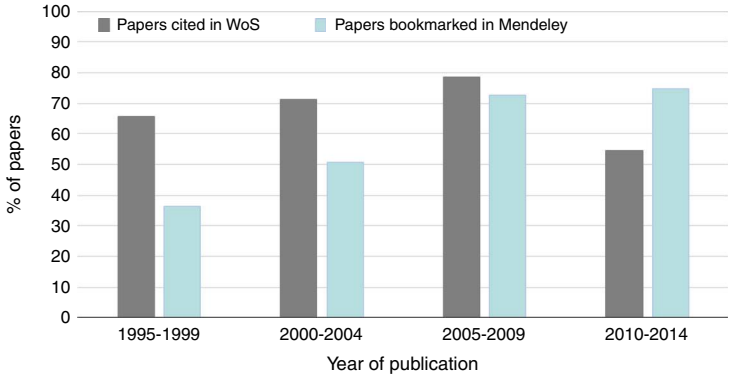

Figure 2.
LIS literature cited and bookmarked by year of publication

the share of cited and bookmarked papers becomes progressively balanced for papers published between 2005 and 2009. In the case of papers published in the last five years (2010-2014), 55 per cent had been cited in Web of Science at the time of data collection, in May 2015, whereas 75 per cent had been bookmarked at least once in Mendeley.

For each paper, Mendeley provides the number of "readers", i.e., the number of users who bookmarked the paper in their libraries. As shown in Table II, one-quarter of the papers (26 per cent) had between one and five users and over half (56 per cent) had between one and 15. At the other end of the scale, an article entitled "Social network sites: definition, history, and scholarship" published in 2007 had 10,217 users. However, this case was particularly extreme, since the second article in the ranking had 893 users.

The distributions of both bookmarks and citations among papers were skewed, since the top 20 per cent of the papers accounted for 75 per cent of the bookmarks and 82 per cent of the citations (Figure 3). However, the top papers by number of bookmarks and by number of citations are not necessarily the same. If we compare the core articles by number of bookmarks (i.e. those with 21 or more bookmarks each) and the core

	Papers	% of papers
1-5 users	8,537	25.6
6-10 users	5,742	17.2
11-15 users	4,380	13.2
16-20 users	3,198	9.6
21-25 users	2,389	7.2
26-30 users	1,701	5.1
31-35 users	1,338	4.0
36-40 users	1,021	3.1
41-45 users	859	2.6
46-50 users	688	2.1
>50 users	3,442	10.3
Total	33,295	100.0

Table II.
Number of users per paper

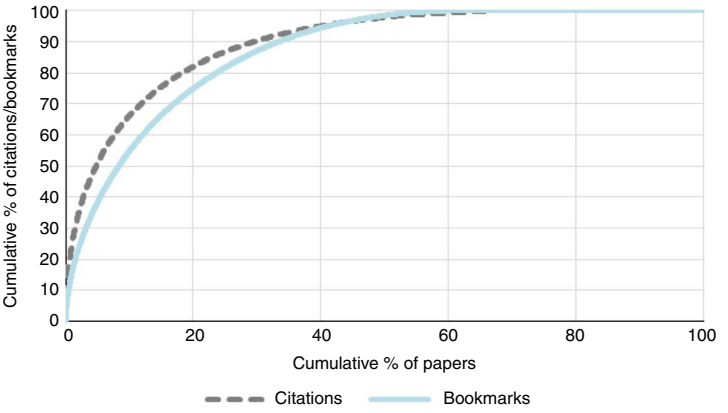

Figure 3.
Distributions of citations and bookmarks

articles by number of citations (i.e. those with ten or more citations each), we observe an overlap of 53 per cent. This means that nearly half of the articles among those with the highest number of bookmarks were not in the top by number citations and vice versa.

3.2 Readership by type of paper (article or review)

In total, 67 per cent of the LIS reviews published between 1995 and 2014 were bookmarked in Mendeley, against 61 per cent of the articles. The median and the average number of bookmarks per review were 2.5 times higher than the number of bookmarks per article (Table III). These differences were similar to those observed regarding citations, with a larger share of reviews being cited (81 per cent) compared to articles (66 per cent). In this case the median and the average number of citations per review were four times the median and the average number of citations per article.

3.3 Readership by language

The number of bookmarks varied according to the language of the paper. As shown in Table IV, the overwhelming majority (95 per cent) of the LIS papers indexed in the Social Sciences Citation Index between 1995 and 2014 were in English. The second language by number of papers was Spanish. Overall, the percentages of English and Spanish papers bookmarked in Mendeley were similar. The number of LIS papers in other languages

	Articles ($n = 53,276$)	Reviews ($n = 1,379$)
Bookmarks		
Papers bookmarked	32,370 (60.8%)	925 (67.1%)
Q1 (bookmarks)	0	0
Median (bookmarks)	4	10
Q3 (bookmarks)	17	39
Average (bookmarks)	13.6	34.0
SD (bookmarks)	34.0	67.0
Citations		
Papers cited	35,349 (66.4%)	1,116 (80.9%)
Q1 (citations)	0	1
Median (citations)	2	8
Q3 (citations)	7	27
Average (citations)	8.2	33.6
SD (citations)	28.6	98.6

Table III. Bookmarks and citations by type of paper (article or review)

Language	Papers	Papers bookmarked	% of papers bookmarked	Average number of bookmarks (SD)	Median number of bookmarks
English	51,912	32,360	62.3	14.7 (53.7)	4
Spanish	1,041	720	69.2	8.6 (15.7)	4
German	716	42	5.9	0.1 (0.4)	0
Portuguese	548	68	12.4	0.6 (3.5)	0
Hungarian	162	21	13.0	0.3 (1.3)	0
French	150	54	36.0	1.2 (2.4)	0
Japanese	118	27	22.9	0.7 (1.7)	0
Other	8	3	37.5	4.1 (9.7)	0
Total	54,655	33,295	60.9	14.2 (52.5)	4

Table IV. Readership by language

indexed in the Social Sciences Citation Index was much smaller and, in all cases, the percentage of papers bookmarked in Mendeley and the average and median number of bookmarks per paper were smaller than among English and Spanish papers.

3.4 Academic status of the users of LIS literature

The largest group of users of LIS literature in Mendeley were PhD students (34 per cent), followed by postgraduate students (22 per cent) and librarians (8 per cent), as shown in Table V.

3.5 Readership by journal

The papers in the sample had been published in 124 journals. Nearly all of them had articles bookmarked, although there were notable differences in the percentage of papers with and without bookmarks (Table VI). In the case of one-fifth (19 per cent) of the journals, nearly all the papers had been bookmarked at least once. However, in the case of one-quarter of other journals, a much smaller share of the papers (less than 20 per cent) had been bookmarked. Table VII details the top 23 journals by percentage of papers bookmarked.

4. Discussion

Studies of readership of academic journals are of interest to the different stakeholders involved in scholarly communication: authors, editors, publishers, librarians, etc. Methods to measure journal readership have traditionally ranged from participative methods such as

Occupation provided by the Mendeley API	Merged	%
PhD student	PhD student	33.8
Doctoral student		
Student (postgraduate)	Postgraduate student	22.2
Student (master)		
Librarian	Librarian	7.5
Assistant professor	Assistant professor	6.9
Lecturer		
Researcher (at an academic institution)	Researcher (at an academic institution)	5.7
Associate professor	Associate professor	5.3
Senior lecturer		
Student (bachelor)	Student (bachelor)	5.0
Professor	Professor	4.2
Postdoc	Postdoc	4.2
Other professional	Other professional	3.3
Researcher (at a non-academic institution)	Researcher (at a non-academic institution)	2.0

Table V.
Users by academic status

% of papers with users	Number of journals	% of journals
91-100	23	18.5
81-90	14	11.3
71-80	14	11.3
61-70	6	4.8
51-60	7	5.6
41-50	8	6.5
31-40	9	7.3
21-30	12	9.7
11-20	15	12.1
1-10	14	11.3
0	2	1.6
Total	124	100

Table VI.
Readership by journal

surveys and interviews, to unobtrusive methods such as shelving statistics, document delivery data, download figures and citation counts. Recently, altmetric indicators have been proposed as a complement to citation metrics for research evaluation since they are available sooner and can capture impact beyond the academic community. In a similar fashion, altmetric indicators such as bookmarks in reference management software might be used to supplement other methods for estimating journal readership.

The results of this study confirm the viability of using bookmark counts to explore the behaviour of users of scholarly literature. The coverage of LIS literature by Mendeley is very extensive. By the first quarter of 2015, Mendeley covered 61 per cent of the LIS literature indexed in the Social Sciences Citation Index in the previous 20 years. This is higher than the percentage observed in previous research, which situated Mendeley's coverage of the social sciences literature somewhere between 47 and 58 per cent (Mohammadi and Thelwall, 2014; Mohammadi *et al.*, 2015). Given that our research analysed a longer period of time than previous studies focussing on more recent literature, the higher coverage observed in this study is particularly significant. Actually, the Mendeley coverage for LIS papers published in the last decade (2005-2014) was higher than 70 per cent. In a similar fashion, although the percentage of LIS articles cited in Web of Science is higher than the percentage of LIS articles bookmarked in Mendeley for

Journal	Papers published	Papers bookmarked	% of papers bookmarked	JCR quartile 2014
Information Technology for Development	111	111	100.0	Q3
Telematics and Informatics	152	151	99.3	Q2
International Journal of Computer-Supported Collaborative Learning	167	165	98.8	Q1
Information Systems Journal	362	354	97.8	Q1
Information & Management	1,097	1,065	97.1	Q1
Library & Information Science Research	464	450	97.0	Q2
Information Systems Research	595	576	96.8	Q1
Information and Organization	88	85	96.6	Q1
Journal of the American Society for Information Science and Technology	2,041	1,958	95.9	Q1
International Journal of Information Management	865	829	95.8	Q1
Information Processing & Management	1,288	1,230	95.5	Q2
Journal of Strategic Information Systems	290	276	95.2	Q1
Knowledge Management Research & Practice	225	213	94.7	Q3
Ethics and Information Technology	151	142	94.0	Q2
International Journal of Geographical Information Science	1,179	1,107	93.9	Q1
Journal of Informetrics	451	422	93.6	Q1
Scientometrics	2,826	2,632	93.1	Q1
Government Information Quarterly	710	661	93.1	Q1
Journal of Knowledge Management	361	335	92.8	Q1
Telecommunications Policy	1,098	1,006	91.6	Q2
Information Development	176	161	91.5	Q3
Serials Review	242	221	91.3	Q3
Journal of Documentation	654	596	91.1	Q2

Table VII.
Top journals by percentage of papers bookmarked

papers published before 2009, the situation is reversed for papers published since 2010. This result confirms that altmetrics offers the important advantage of speed compared to citations for estimating readership.

In addition to speed, another possible advantage of altmetric indicators over citation counts is that they can help to estimate impact outside the academic community. This is especially interesting in fields such as LIS where practicing professionals account for nearly a quarter of the literature published worldwide (Walters and Wilders, 2016). Our results show that, for LIS scholarly literature, librarians and other professionals account for more than 10 per cent of the bookmarks, even though they still come behind PhD and postgraduate students. The number of librarians who use the tool to manage professional literature may also help to explain the higher coverage of LIS literature in Mendeley compared to other disciplines in other social sciences, since in other disciplines most readers are academics – mainly students but also faculty and researchers (Mohammadi *et al.*, 2015).

The distribution of both bookmarks and citations among papers is skewed, but the rankings of papers by number of bookmarks and by number of citations are only partly coincidental. That means that there are articles which have a high number of bookmarks but are not among the top papers by number of citations and vice versa. The reasons for these discrepancies may include (in addition to the different life-cycles of bookmarks and citations) the presence of articles that attract communities which use research but do not cite it, such as students and professionals (Thelwall, 2016).

The results also show that the share of reviews bookmarked in Mendeley is higher than that of articles. Additionally, the average and median number of readers per document is higher for reviews than for articles. This is consistent with previous studies that show that

literature reviews are more often cited than regular articles, most likely due to their breadth (Teixeira *et al.*, 2013).

English is the lingua franca of science, and LIS is no exception to this rule. More than 95 per cent of the LIS literature indexed in the Social Sciences Citation Index is in English. In terms of readership, papers in English are more often bookmarked in Mendeley than papers in any other language. Additionally, there are clear differences in the bookmarking of LIS journals. The share of bookmarked articles varies dramatically from journal to journal: some journals have no articles bookmarked at all, while others have nearly all articles bookmarked by at least one reader.

5. Conclusions

The results of this study show that altmetrics can be used as a source of data in information behaviour studies. Reference management software provides an unobtrusive means of capturing reading habits in scholarly literature that are useful to all the stakeholders in the scholarly communication system.

The application of altmetric indicators to supplement citations counts in order to estimate readership presents two advantages over the use of citations alone. Bookmarks are available sooner, as shown by the fact that the percentage of recent literature bookmarked in Mendeley is much higher than the share of literature cited. Additionally, bookmarks are useful to capture usage beyond the academic community, since reference management software can be employed by professionals to manage the literature. This is especially relevant in fields such as LIS where practicing professionals account for a large part of the use of the literature.

References

Bonasio, A. (2013), "Mendeley has 2.5 million users!", *Mendeley Blog*, available at: https://blog.mendeley.com/2013/09/03/mendeley-has-2-5-million-users/ (accessed 1 June 2016).

Bornmann, L. (2014), "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics", [redacted], Vol. 8 No. 4, pp. 895-903.

Bornmann, L. and Haunschild, R. (2015), "Which people use which scientific papers? An evaluation of data from F1000 and Mendeley", [redacted], Vol. 9 No. 3, pp. 477-487.

Borrego, A. and Fry, J. (2012), "Measuring researchers' use of scholarly information through social bookmarking data: a case study of BibSonomy", [redacted], Vol. 38 No. 3, pp. 297-308.

Haustein, S. (2014), "Readership metrics", in Cronin, B. and Sugimoto, C.R. (Eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press, Cambridge, pp. 327-344.

Haustein, S. and Siebenlist, T. (2011), "Applying social bookmarking data to evaluate journal usage", [redacted], Vol. 5 No. 3, pp. 446-457, available at: http://dx.doi.org/10.1016/j.joi.2011.04.002

Jamali, H.R., Nicholas, D., Watkinson, A., Herman, E., Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R. and Nichols, F. (2014), "How scholars implement trust in their reading, citing and publishing activities: geographical differences", [redacted], Vol. 36 Nos 3-4, pp. 192-202.

Li, X., Thelwall, M. and Giustini, D. (2011), "Validating online reference managers for scholarly impact measurement", [redacted], Vol. 91 No. 2, pp. 461-471.

Maflahi, N. and Thelwall, M. (2016), "When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals", [redacted], Vol. 67 No. 1, pp. 191-199.

Mohammadi, E. and Thelwall, M. (2014), "Mendeley readership altmetrics for the social sciences and humanities: research evaluation and knowledge flows", [redacted], Vol. 65 No. 8, pp. 1627-1638.

Mohammadi, E., Thelwall, M. and Kousha, K. (2016), "Can mendeley bookmarks reflect readership? A survey of user motivations", , Vol. 67 No. 5, pp. 1198-1209.

Mohammadi, E., Thelwall, M., Haustein, S. and Larivière, V. (2015), "Who reads research articles? An altmetrics analysis of Mendeley user categories", , Vol. 66 No. 9, pp. 1832-1846.

Teixeira, M.C., Thomaz, S.M., Michelan, T.S., Mormul, R.P., Meurer, T., Fasolli, J.V. and Silveira, M.J. (2013), "Incorrect citations give unfair credit to review authors in ecology journals", *PLoS ONE*, Vol. 8 No. 12, p. e81871, available at: http://dx.doi.org/10.1371/journal.pone.0081871

Thelwall, M. (2009), *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*, , San Rafael, CA.

Thelwall, M. (2016), "Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa?", *Journal of Librarianship and Information Science*, available at: http://doi.org/10.1177/0961000615594867

Thelwall, M. and Sud, P. (2016), "Mendeley readership counts: an investigation of temporal and disciplinary differences", , Vol. 67 No. 12, pp. 3036-3050.

Thomson Reuters (2013), "Web of science help: searching the document type field", available at: http://images.webofknowledge.com/WOKRS59B4/help/WOS/hs_document_type.html (accessed 1 June 2016).

Van Noorden, R. (2014), "Online collaboration: scientists and the social network", , Vol. 512 No. 7513, pp. 126-129.

Walters, W. and Wilders, E. (2016), "Disciplinary, national, and departmental contributions to the literature of library and information science, 2007-2012", , Vol. 67 No. 6, pp. 1487-1506.

Zahedi, Z., Costas, R. and Wouters, P. (2014), "How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications", , Vol. 101 No. 2, pp. 1491-1513.

About the authors

Aida Pooladian is a PhD Student in the Faculty of Library and Information Science at the University of Barcelona (Spain). Her research interests are altmetrics, scientometrics, social media, social networks and webometrics.

Ángel Borrego is a Lecturer in the Department of Library and Information Science at the University of Barcelona (Spain). He teaches courses in Scientific Information and Research Methods. His research interests focus on scholarly communication and scientists' information behaviour. Ángel Borrego is the corresponding author and can be contacted at: borrego@ub.edu

Journal of Informetrics 10 (2016) 1135–1142

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

A longitudinal study of the bookmarking of library and information science literature in Mendeley

Aida Pooladian, Ángel Borrego*

Universitat de Barcelona, Facultat de Biblioteconomia i Documentació, Melcior de Palau, 140, 08014 Barcelona, Spain

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form 6 October 2016
Accepted 6 October 2016
Available online 18 October 2016

Keywords:
Altmetrics
Citations
Library and information science
Mendeley

ABSTRACT

Speed and breadth have been suggested as two advantages of altmetrics over citation counts since they might estimate impact immediately after publication and beyond the academic community of authors. In order to investigate the validity of these claims, we performed a fifteen-month longitudinal study of the evolution of bookmarks in Mendeley for a set of 3813 articles published in Library and Information Science in 2014. Results show that 87.6% of the literature was bookmarked at least once by May 2016 whereas only 55% was cited. The correlation between bookmarks and citations was moderate and the overlap between the most frequently bookmarked and the most frequently cited papers increased over time. A significant share of the bookmarks were made by students and professionals, although the shares of bookmarks made by different categories of users changed as time went by. Bookmarks made by users based in less wealthy nations also increased over time. The study is limited by the incomplete information provided by Mendeley regarding users' academic status and country of residence, the upgrades of the software used in data collection, and the fact that one year is a rather long publication period for a longitudinal study of a fast-changing feature like bookmarks.

1. Introduction

Longitudinal studies of article age measure the obsolescence or decay of scientific information. The rationale is that the usage of scientific literature declines as time goes by and new research findings render previous results obsolete. Although some studies on article age have applied shelving statistics or document delivery data to investigate obsolescence, Nicholas et al. (2005) observed that most studies are conducted using citation analysis. In the print era, obsolescence research was useful to manage library collections, identifying opportunities to remove low-use materials in order to accommodate new collections. In the current digital era of access to electronic journals, authors have wondered whether the use of older materials increases thanks to easier online access. So far, the research has yielded contradictory results. Whereas Odlyzko (2002) stated that easy access to journal archives might increase the use of older materials, Tenopir and King (2000) and Nicholas et al. (2005) observed that most of the use of an article occurs in the first year after its publication. However, Martín-Martín, Orduña-Malea, Ayllón and Delgado López-Cózar (2016) recently revealed an elevated and growing percentage of citations of articles at least 10 years old.

* Corresponding author.
E-mail addresses: aida.pouladian@gmail.com (A. Pooladian), borrego@ub.edu (Á. Borrego).

http://dx.doi.org/10.1016/j.joi.2016.10.003
1751-1577/

In recent years, altmetrics have been proposed as a complement to traditional citation counts for research evaluation. Altmetric indicators estimate research impact by quantifying the dissemination of scholarly output in social media. Examples include mentions in blogs, number of tweets and retweets or inclusion in social bookmarking services. Several studies have measured the breadth of coverage of altmetric sources, taking citation databases such as Web of Science (WoS) and Scopus as points of comparison. Most research concurs that Mendeley, a reference management software, is the altmetric data source with the most extensive coverage of scientific literature (Mohammadi & Thelwall, 2014; Zahedi, Costas, & Wouters, 2014; Mohammadi, Thelwall, Haustein, & Larivière, 2015).

Altmetrics could help to overcome two of the most frequent criticisms levelled at citation analysis (Bornmann, 2014). First, altmetrics can estimate impact immediately after publication, whereas citations need much more time to accrue. The accumulation of citations is a slow process that usually takes years. As a result, research evaluation does not usually focus on the citations received by individual articles, but rather on the impact of the journal in which the article was published. By contrast, altmetric indicators of research visibility at the article level can be calculated almost immediately since an article can be blogged, tweeted or bookmarked directly after publication. Second, altmetrics can estimate impact among non-academic segments of the audience, such as professionals who use research results in their everyday practice, whereas citations only estimate impact within the academic community of authors who write, and hence cite.

This paper aims to investigate the validity of these two potential advantages of altmetrics, namely speed and breadth. In order to do so, we conducted a fifteen-month longitudinal study of the evolution of the bookmarks in Mendeley for a set of articles published in Library and Information Science (LIS) in 2014. A comparison between bookmarks and citations and an analysis of the academic status of Mendeley users allowed us to investigate whether bookmarks are useful to estimate impact at shorter notice than citations and beyond the academic community. The literature in LIS was chosen as a case study because the topic is well-known by the authors of the paper, making interpretation of the results easier. Additionally, it is a discipline where practising professionals are active in research tasks, accounting for nearly a quarter of all LIS literature published worldwide (Walters & Wilders, 2016).

As far as we know, this is the first longitudinal study of the evolution of Mendeley data that looks at the increase in bookmarks for a set of articles over time. Specifically, the study is underpinned by the following research questions:

- What is the evolution in the coverage of LIS literature in Mendeley compared to the evolution in the number of citations received in WoS?
- What is the evolution in the population of users of LIS literature by academic status?
- What is the evolution in the population of users of LIS literature by country of residence?
- Are there any differences in the bookmarking of articles and reviews?

2. Methods

According to Moed (2005), longitudinal studies can be conducted using either a synchronous or a diachronous approach. In the former, the researcher analyses the number of citations or downloads (in our case, bookmarks) as a function of the papers' publication date, which is fixed. In the latter, the researcher measures the citations or downloads (or bookmarks) for a set of papers as a function of time.

In order to conduct a synchronous analysis, the number of bookmarks for a set of papers published in a given year were recorded each month for fifteen months. In February 2015, we used the Social Sciences Citation Index to retrieve a total of 3813 papers published in 2014 in the category "Information Science & Library Sciences". The search was limited to articles (3750 papers) and reviews (63 papers), and our analysis excluded all other types of document, such as book reviews, editorials, letters or proceedings. Throughout this article we use the term "papers" to refer only to this set of articles and reviews.

From March 2015 to May 2016, each paper was searched monthly in Mendeley using *Webometric Analyst* software (Thelwall, 2009) using a query containing the title of the article, the first author's last name, the year of publication and the Digital Object Identifier (DOI). For each paper, the number of users, their academic status and their country of residence were recorded from Mendeley on the first Thursday of each month. At this point, it should be borne in mind that Mendeley only returns the three main academic status categories and the three main countries of residence for the users of each paper.

Mendeley's fifteen academic status categories were merged into six broader categories in order to facilitate the analysis (Table 1). The original categories provided by Mendeley overlapped and were difficult to distinguish, since users are obliged to select one category but are not provided with thorough definitions. Actually, since the study was conducted, Mendeley itself has grouped some categories, so the options offered now to a new user of the platform do not correspond to those available at the time of data collection. The breakdown of users by academic status in the last month considered, May 2016, seemed to be erroneous, since the results were identical to those in April 2016. Therefore, the results for May 2016 were removed from the analysis of users' academic status. Regarding the users' country of residence, the analysis compared the United States (US), the 28 member states of the European Union (EU-28) and 90 countries with a gross domestic product per capita based on purchasing power parity (PPP GDP) below 25,000 current international dollars (World Bank, 2016).

The papers' publication date was set as 2014. While this was the year of publication of the papers' version of record, some of the earliest articles may already have been published prior to the end of 2013. Similarly, some articles dated as 2014 could have been published in early 2015. This is a limitation for a longitudinal study of a fast-changing feature like bookmarks.

Table 1
Mendeley user categories by academic status.

Mendeley categories	Merged
PhD student	PhD student
Doctoral student	
Postdoc	Postdoc
Student bachelor	Student
Student master	
Postgraduate	
Professor	Faculty
Senior lecturer	
Assistant professor	
Associate professor	
Lecturer	
Researcher academic	Researcher
Researcher non academic	
Librarian	Professional
Other professional	

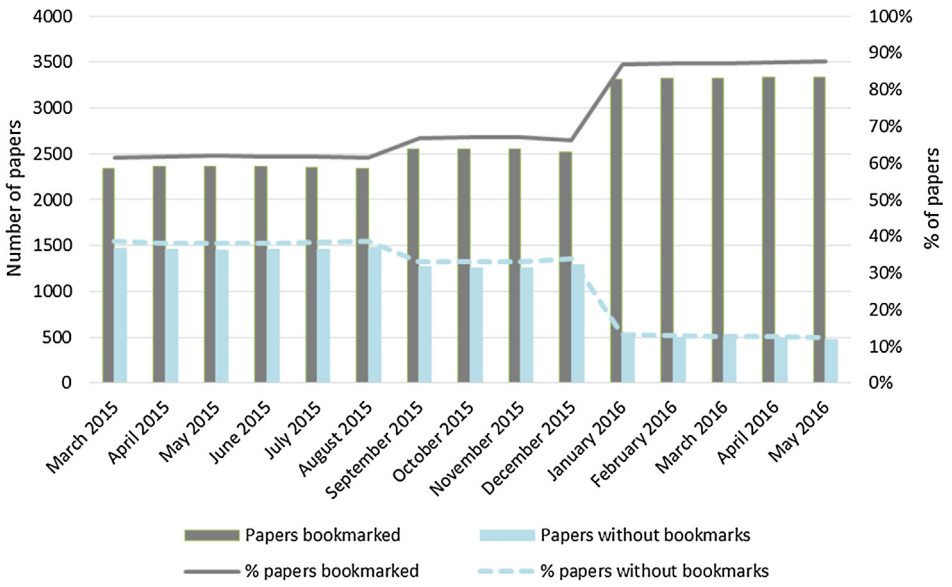

Fig. 1. Evolution in the coverage of LIS literature bookmarked in Mendeley.

The number of times each paper had been cited was obtained from the WoS twice: at the beginning of the study period, in February 2015, and at the end, in May 2016. In order to supplement the synchronous analysis with a diachronous approach, the number of Mendeley bookmarks for a set of 54,655 papers published in LIS between 1995 and 2014 were retrieved in May 2015 from the Web of Science. Each record was searched in Mendeley in a similar fashion to that used in this study. More information about this dataset is available in Pooladian and Borrego (2016).

3. Results

3.1. Evolution in the coverage of LIS literature in Mendeley

Out of the 3813 LIS papers published in 2014, 2342 (61.4%) had been saved in Mendeley by at least one user by the beginning of March 2015 (Fig. 1). Fifteen months later, in May 2016, the coverage had risen to 87.6%, i.e., 3341 papers had been bookmarked at least once.

While Fig. 1 shows two peaks in the share of articles bookmarked—a soft one in September 2015 and a sharper one in January 2016—it is unclear whether these increases reflect a change in users' bookmarking behaviour or are caused by changes in the software used for data collection, which was upgraded several times during the study period. In particular, a few odd cases were observed in the monthly evolution in the number of bookmarks. Despite an increase in the number of bookmarks and the percentage of articles bookmarked from August to September 2015, 179 papers that had 10 or more bookmarks experienced a decrease between 50% and 100% in the number of bookmarks over that period. Most of these papers fell to zero bookmarks and then recovered them in January 2016, a situation that helps to explain the increase observed in

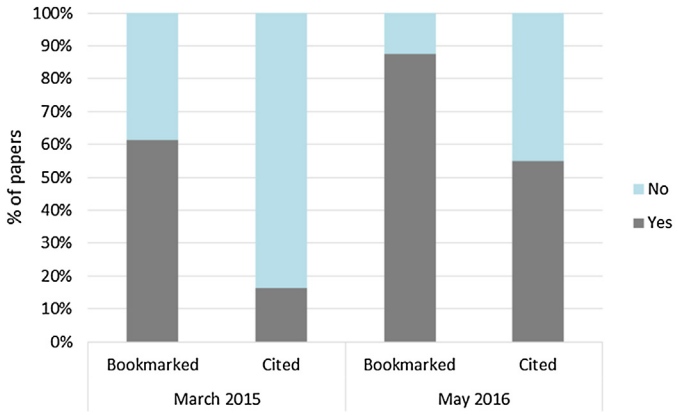

Fig. 2. LIS literature bookmarked and cited.

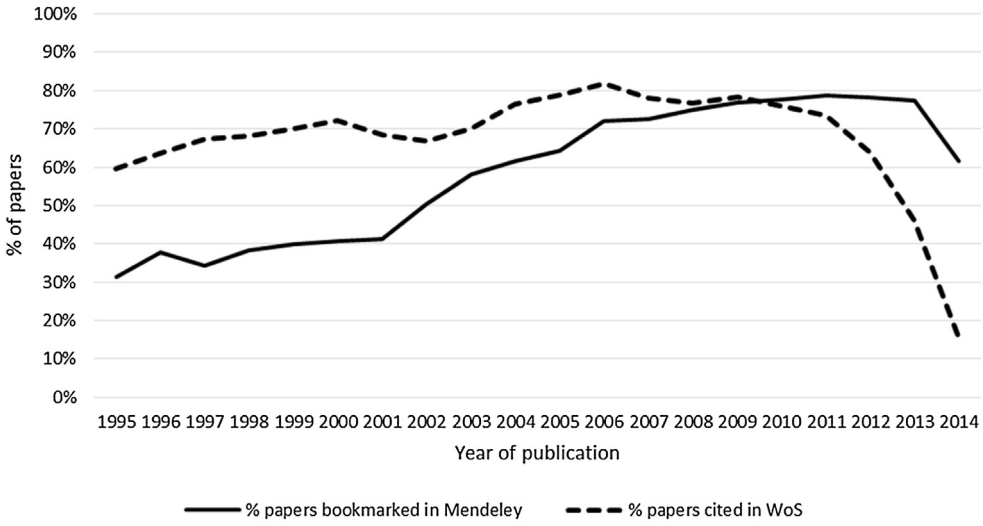

Fig. 3. Evolution in the coverage of LIS literature bookmarked in Mendeley and cited in WoS by year of publication (n = 54,655 articles, data collected in March 2015).

the latter month. Two additional situations were observed in which papers with more than 10 bookmarks lost more than 50% of them from one month to the next, affecting 21 and 26 papers, respectively.

A potential advantage of bookmarks over citations is speed, since articles can be bookmarked immediately after publication, whereas citations need much more time to accrue. Fig. 2 shows that by March 2015, 61.4% of the LIS papers published in 2014 had been bookmarked at least once in Mendeley, whereas only 16.4% had been cited in WoS. At the end of the study period, in May 2016, 87.6% of the LIS papers published in 2014 had been bookmarked, whereas only 55.0% had received at least one citation.

Fig. 3 supplements the synchronous approach with a diachronous analysis of the number of bookmarks and citations in May 2015 for a set of 54,665 LIS papers published between 1995 and 2014. The results show that the percentage of cited literature is higher for papers published up to 2009, i.e., six years prior to data collection. For recent papers, i.e., those published in the last six years, the share of literature bookmarked is higher than that of literature cited. Although the shares of literature bookmarked and cited both decline for recent papers, the decline is much more acute for citations than for bookmarks. When interpreting these results, it must be borne in mind that the first public beta version of Mendeley was released in August 2008.

Table 2 details the bookmarking data for the set of papers. In the fifteen-month study period the number of bookmarks multiplied fourfold, from 19,563 to 83,845, whereas citations multiplied by six, from 993 to 6515. Not only did the median and the average number of bookmarks per paper increase, but so did the scattering of the bookmarks across the dataset. Thus, as Fig. 4 shows, the top 20% of the most frequently bookmarked articles in March 2015 accounted for 70% of the bookmarks. Fifteen months later, the set of 20% most frequently bookmarked articles accounted for 60% of the bookmarks.

The correlation between bookmarks and citations was moderate throughout the study period, rising slightly from Spearman's rho = 0.52 in March 2015 to 0.56 in May 2016. When comparing the rankings of top papers by number of bookmarks

Table 2
Evolution in readership of LIS papers in Mendeley.

	3/15	4/15	5/15	6/15	7/15	8/15	9/15	10/15	11/15	12/15	1/16	2/16	3/16	4/16	5/16
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q1	0	0	0	0	0	0	0	0	0	0	4	4	4	4	5
Med	2	2	2	3	3	4	6	6	7	7	11	11	11	12	13
Q3	7	8	9	10	11	14	15	16	18	19	24	25	25	26	28
Max	94	118	129	145	153	164	207	215	234	262	245	247	253	268	291
Average	5.3	5.9	6.7	7.6	8.1	10.1	12.0	12.5	13.5	14.6	18.6	19.0	19.4	20.6	22.0
SD	8.9	10.1	11.3	12.4	13.2	16.4	18.3	18.9	20.3	21.9	24.2	24.7	25.5	27.1	28.7
Bookmarks	19563	22061	25110	28166	30203	37573	44547	46457	50090	54274	69395	71112	73816	78621	83845
Citations	993	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	6515

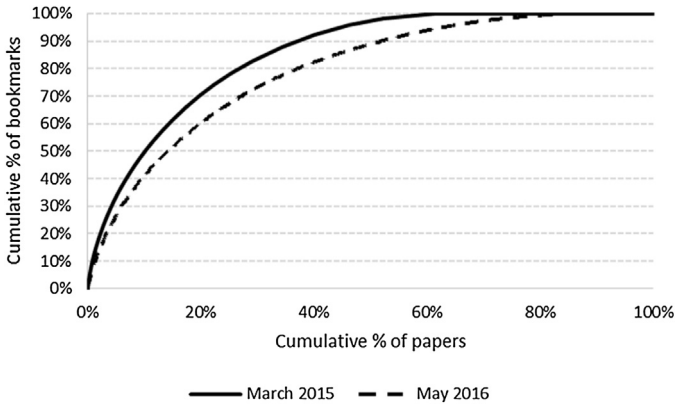

Fig. 4. Scattering of bookmarks across the LIS literature.

and number of citations, we observe significant differences. Specifically, in March 2015 there were 94 papers in the top 20% of articles by bookmarks (they had 30 or more bookmarks each). However, only 15 of these papers (16%) were simultaneously in the top 20% of articles by number of citations (they had received four or more citations each). By May 2016 the overlap had increased to 30%. That is, 41 papers out of the 128 in the top 20% of papers ranked by bookmarks (93 or more bookmarks each) were also in the top 20% of papers by citations (i.e., they had received ten or more citations each).

3.2. Bookmarks by users' academic status

Fig. 5 shows the evolution in the share of bookmarks by users' academic status. The share of papers bookmarked by faculty increased sharply during the study period, rising from 13% of the bookmarks in March 2015 to 38% in the last few months analysed. The opposite trend is observed among students, who were responsible for 30.7% of the bookmarks in March 2015, a percentage that fell to 14.3% at the end of the study period. Similarly, the share of bookmarks by professionals and researchers gradually declined, whereas that of PhD students remained fairly stable at around 30% throughout the whole study period.

3.3. Bookmarks by users' country of residence

LIS articles were bookmarked by users based in 165 countries. In order to facilitate the analysis, Fig. 6 presents the evolution in the share of bookmarks made by users in the United States (US), in the 28 member states of the European Union (EU-28) and in 90 countries with a gross domestic product per capita based on purchasing power parity below 25,000 current international dollars (PPP GDP < $25,000). All together, these three categories of country represent approximately 80% of the bookmarks in the dataset.

Results: show that the share of bookmarks by US users fell from 22.5% at the beginning of the study period to 19.0% at the end, whereas the share in countries with a lower GDP increased from 17.8% to 24.6%. The percentage of bookmarks made by users based in Europe remained stable at around 40% throughout the whole study period.

3.4. Bookmarks of articles and reviews

Fig. 7 shows that reviews were bookmarked more frequently than articles. At the beginning of the study period, in March 2015, 61.2% of the articles had been bookmarked at least once in Mendeley compared to 76.2% of the reviews. Fifteen months

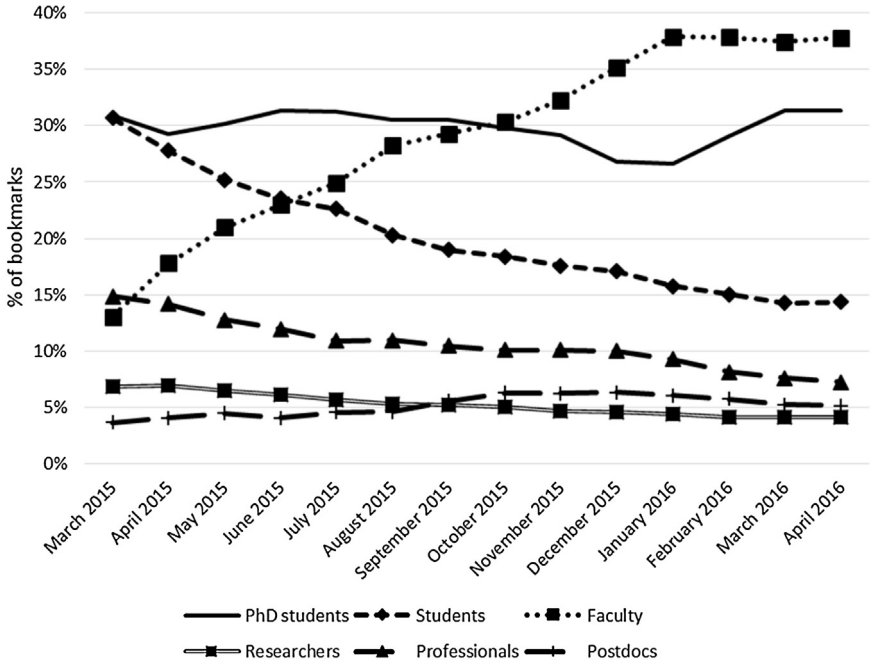

Fig. 5. Evolution in the share of bookmarks by users' academic status.

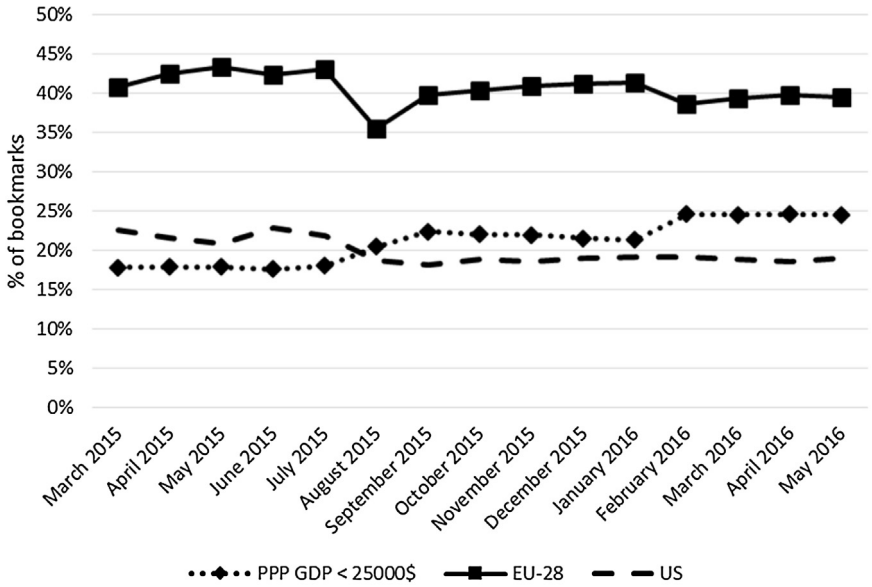

Fig. 6. Evolution in the share of bookmarks by users' country of residence.

later, the percentages had risen to 87.5% of the articles and 96.8% of the reviews. The median number of bookmarks per review was about 2.5 times higher than the median number of bookmarks per article throughout the whole study period.

4. Discussion and conclusions

This study provides evidence to support the validity of two suggested advantages of bookmarks over citations in estimating the impact of scientific literature. Papers are bookmarked immediately after publication, whereas citations need much more time to accrue. Thus, nearly nine of every ten LIS papers are bookmarked in Mendeley in the year following publication, a figure that greatly exceeds that of articles cited. This result is consistent with results observed in the only two previous studies to investigate the role of time in the evolution of Mendeley bookmarks. Specifically, Maflahi and Thelwall (2016), using a diachronous approach, observed that, for a set of four LIS journals, papers initially attracted more Mendeley

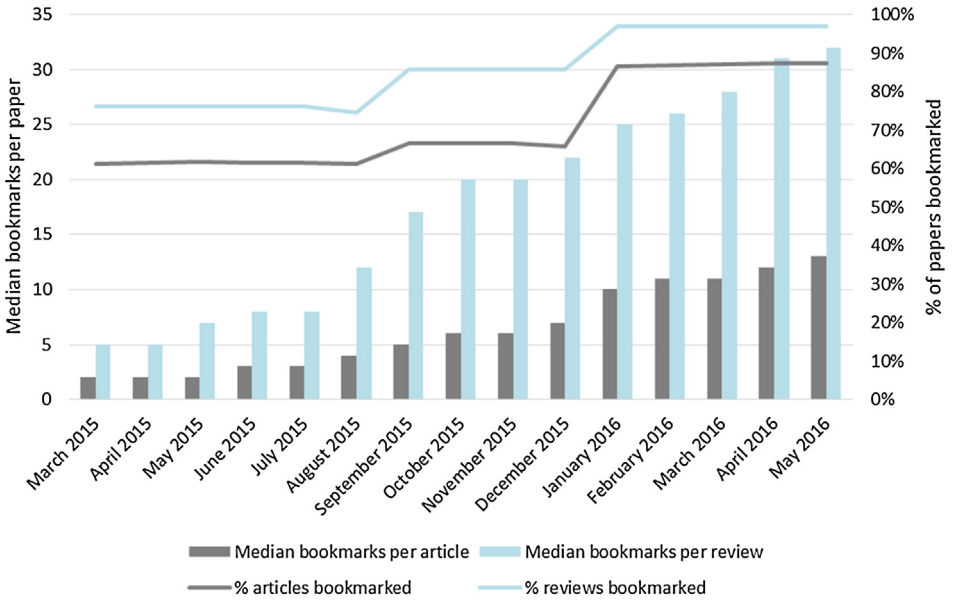

Fig. 7. Evolution of the bookmarks of articles and reviews.

bookmarks than Scopus citations, but the situation reversed after approximately seven years. A further study expanded the analysis to five disciplines and confirmed that the correlation between bookmarks and citations tended to increase over five years and then stabilise (Thelwall & Sud, 2016).

Similarly, the results are consistent with those observed in studies on the usage of scholarly literature, such as those by Tenopir and King (2000), who showed that nearly two-thirds of all use occurs in an article's first year of publication. Nicholas et al. (2005) reached a similar conclusion by analysing download data that showed that about 55% of use is for items published within the previous 15 months. Schlögl, Gorraiz, Gumpenberger, Jack and Kraker (2014) also showed that the highest number of downloads of papers are usually made in the publication year and immediately afterwards. All in all, these results suggest that Mendeley bookmarks can be a good complement to citations in the assessment of young articles, provided they are in a discipline with a high level of Mendeley users. According to our results, this is the case of LIS where Mendeley coverage of the literature is higher than in other Social Sciences where the coverage is somewhere between 47% and 58% (Mohammadi & Thelwall, 2014; Mohammadi et al., 2015).

The core sets of papers by number of bookmarks and number of citations only partly coincide. This means that some articles having a high number of bookmarks are not among the top papers by number of citations and vice versa. In addition to the different life-cycles of bookmarks and citations, the reasons for this discrepancy may include the presence of articles that attract communities which use research but do not cite it, such as students and professionals (Thelwall, 2016). This evidence supports the second potential advantage of bookmarks over citations, i.e., that bookmarks reflect impact beyond the academic community of authors. Our results show that a significant share of the bookmarks of LIS literature are made by students and professionals who are not necessarily authors and therefore do not cite. At this stage some unexpected results were observed in the evolution in the share of papers bookmarked by the different categories of users. While the amount of bookmarks increased across all categories of users, the increase in the share of bookmarks made by faculty was significantly higher. This suggests that faculty do not bookmark papers immediately after publication, but do tend to bookmark papers over time. Another explanation might be an increase in the population of faculty users who employ Mendeley. Further research will be necessary to interpret this result. A tentative analysis was performed in order to compare the journals bookmarked by different categories of users in terms of their orientation (computer science, librarianship, management, etc.) but no significant results were obtained and the analysis was discarded.

Access to scholarly information is a major concern in developing countries where fewer resources are available for journal subscriptions. The high cost of Western scientific journals poses a major barrier to researchers in developing nations (Davies & Walters, 2011). This situation may be mirrored in the evolution in the share of bookmarks by users' country of residence. Thus, a reduction in the share of articles bookmarked by users based in the US is compensated by an increase in the share of bookmarks in less wealthy nations with a lower GDP. With the passage of time, open-access versions of articles become available in repositories and other websites that may stimulate an increase in the amount of bookmarks. Actually, the differences in the evolution of the shares of bookmarks by country of residence could be sharper than suggested by the results, since Mendeley only returns the three main countries of residence for the users of each paper.

The results also show that the share of reviews bookmarked in Mendeley is higher than that of articles. Additionally, the average and median number of users per document is higher for reviews than for articles. This is consistent with previous

studies that show that literature reviews are more often cited than regular articles, most likely due to their breadth (Schlögl et al., 2014; Teixeira et al., 2013).

In addition to the fact that Mendeley only returns the three main academic status categories and the three main countries of residence for the users of each paper, two further limitations need to be borne in mind when interpreting the results. First, upgrades in the software employed for data collection may have affected the reliability of the data as reflected by fluctuations in the monthly amount of bookmarks for some papers. Second, the papers' publication date was set as 2014, but this is only the year of publication of the version of record and their dates of online availability might well precede or extend beyond that year.

Authors' contributions

Conception and design of analysis: Aida Pooladian, Ángel Borrego.
Data collection: Aida Pooladian.
Analysis: Aida Pooladian, Ángel Borrego.
Paper: Ángel Borrego.

References

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*(4), 895–903.

Davies, P. M., & Walters, W. H. (2011). The impact of free access to the scientific literature: A review of recent research. *Journal of the Medical Library Association*, *99*(3), 208–217.

Maflahi, N., & Thelwall, M. (2016). When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals. *Journal of the Association for Information Science and Technology*, *67*(1), 191–199.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & Delgado López-Cózar, E. (2016). Back to the past: On the shoulders of an academic search engine giant. *Scientometrics*, *107*(3), 1477–1487.

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, *56*(10), 1088–1097.

Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, *65*(8), 1627–1638.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, *66*(9), 1832–1846.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali, H. R., & Polydoratou, P. (2005). Revisiting 'obsolescence' and journal article 'decay' through usage data: An analysis of digital journal use by year of publication. *Information Processing and Management*, *41*(6), 1441–1461.

Odlyzko, A. M. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, *15*(1), 7–19.

Pooladian, A., & Borrego, A. (2016). Twenty years of readership of library and information science literature under Mendeley's microscope. *Performance Measurement and Metrics*, *17*(3), in press.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics*, *101*(2), 1113–1128.

Teixeira, M. C., Thomaz, S. M., Michelan, T. S., Mormul, R. P., Meurer, T., Fasolli, J. V., et al. (2013). Incorrect citations give unfair credit to review authors in Ecology journals. *PLoS One*, *8*(12), e81871.

Tenopir, C., & King, D. W. (2000). *Towards electronic journals: Realities for scientists, librarians, and publishers*. Washington, DC: Special Libraries Association.

Thelwall, M., & Sud, P. (2016). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology* [in press].

Thelwall, M. (2009). *Introduction: To webometrics: Quantitative web research for the Social Sciences*. San Rafael, CA: Morgan & Claypool.

Thelwall, M. (2016). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? *Journal of Librarianship and Information Science* [in press].

Walters, W., & Wilders, E. (2016). Disciplinary, national, and departmental contributions to the literature of Library and Information Science, 2007–2012. *Journal of the Association for Information Science and Technology*, *67*(6), 1487–1506.

World Bank. (2016). *World development indicators database: GDP per capita, PPP.*. http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?order=wbapi_data_value_2014+wbapi_data_value+wbapi_data_value-last&sort=desc. (Accessed 14.05.16)

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, *101*(2), 1491–1513.

Scientometrics
DOI 10.1007/s11192-017-2474-z

Methodological issues in measuring citations in Wikipedia: a case study in Library and Information Science

Aida Pooladian[1] · **Ángel Borrego**[1]

Received: 15 May 2017

Abstract Wikipedia citations have been suggested as a metric that partially captures the impact of research, providing an indication of the transfer of scholarly output to a wider audience beyond the academic community. In this article, we explore the coverage of Library and Information Science literature published between 2001 and 2010 in Wikipedia, paying special attention to the methodological issues involved in counting Wikipedia citations. The results reveal severe limitations in the use of Wikipedia citations for research evaluation. Lack of standardization and incompleteness of Wikipedia references make it difficult to retrieve them. The number of Wikipedia citations is very low, with less than 3% of articles in the sample having been cited. A significant number of references are cited in biographical entries about the authors of the articles, resulting in a phenomenon of accumulated advantage, which is similar to the Matthew effect. Nearly one-third of the Wikipedia citations link to an open access source, although this result is probably an underestimate of open access availability, given the incompleteness of Wikipedia citations.

Keywords Altmetrics · Citation analysis · Library and Information Science · Wikipedia

✉ Ángel Borrego
borrego@ub.edu

Aida Pooladian
aida.pouladian@gmail.com

[1] Facultat de Biblioteconomia i Documentació, Universitat de Barcelona, Melcior de Palau 140, 08014 Barcelona, Spain

Published online: 27 July 2017

Introduction

Wikipedia, a blend of the words "wiki"—a technology that allows collaborative modification of a website—and "encyclopedia", is a free online encyclopedia, written collaboratively by the people who use it. At present, Wikipedia is the largest online encyclopedia. It held nearly 5.4 million articles in the English version alone in May 2017.[1]

Altmetrics are non-traditional metrics proposed as an alternative or a complement to traditional citation impact metrics. Altmetrics cover other aspects of the impact of scientific works, such as the number of views, downloads, bookmarks or mentions in social media. A reference to a scientific article in Wikipedia can be seen as a metric that partially captures the impact of the article. Contrary to other sources of altmetric data, such as social media, in which the easiness of the process may result in casual sharing of research results, citations in Wikipedia may be indicative of stronger engagement of the user with the article. Among its "five pillars", Wikipedia enforces strict editorial guidelines striving "for verifiable accuracy, citing reliable, authoritative sources" that ensure quality and standard across all the encyclopedia entries.[2] Citations allow Wikipedia editors to make their contributions verifiable by supporting them with trustworthy external sources, and enable readers to locate further information on topics of interest. Thus, citations in Wikipedia can be considered an indication of the transfer of scholarly output to a wider audience.

Nielsen (2007) was one of the first authors to examine citations in Wikipedia to articles in scholarly journals. He observed that Wikipedia citations correlated strongly with the total number of citations to a journal, but more weakly with the journal's impact factor. Wikipedia contributors also showed a slight tendency to cite articles in high-impact journals such as *Nature* and *Science*. A similar trend was described by Stankus and Spiegel (2010), who observed that both titles topped the list of Wikipedia journal sources for entries on the brain and behavioural sciences. However, the results are different in disciplines with distinct citing behaviours. Thus, Luyt and Tan (2010) found that most citations in a set of Wikipedia history entries were to books, with very few citations of academic journal material. Similarly, Halfaker and Taraborelli (2015) analysed the presence of ISBN, PubMed, DOI and arXiv identifiers in Wikipedia and found that most matches were to books and monographs. To sum up, citations in Wikipedia of scholarly literature have been used as proxy measurements of the encyclopedia's reliability, and differences in verifiability across topics have been identified (Mesgari et al. 2015).

Using a different approach, Huvila (2010) conducted a survey on Wikipedia editors' information behaviour, identifying five groups of contributors who use different information sources. The results indicated a preference among contributors for sources that are available online, although a significant proportion of the original information was based on printed literature, personal expertise and other non-digital sources of information.

Finally, another line of inquiry has explored Wikipedia as an alternative source of evidence about the impact of research. Thus, Evans and Krauthammer (2011) searched PubMed IDs and DOIs in Wikipedia and observed that these articles have higher citation counts than an equivalent random article subset. The fact that articles were cited in Wikipedia soon after publication suggested that Wikipedia citations might represent a resource for assessing articles' impact. This opinion was shared by one-third of the bibliometricians who attended the 17*th International Conference on Science and Technology Indicators* (STI2012), who believed that the number of Wikipedia links or mentions of an

[1] https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia.

[2] https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

article could be of use in author or article evaluation (Haustein et al. 2014). Using a different approach, Tarango et al. (2017) analysed the obsolescence of Wikipedia featured articles in Spanish and observed that more than 90% had last been modified in the two years previous to data collection.

Interest in Wikipedia as a source of altmetric data has grown in recent years. In February 2015, Altmetric.com, a start-up focused on tracking and analysing online activity relating to scholarly literature, announced that any mentions of articles and academic output in Wikipedia would be reflected in a new Wikipedia tab on the Altmetric details page.[3] In order to capture this information, the academic output that was mentioned had to be referenced with proper Wikipedia citation tags.[4] However, exploratory research led to doubts about the use of Wikipedia as a source of evidence of the impact of research. Lin and Fenner (2014) found that just 4% of PLOS articles had been cited in Wikipedia. Thelwall (2016) analysed the presence of astronomy and astrophysics research in Wikipedia, and indicated that the use of Wikipedia citations as a proxy for public interest in research articles was limited, due to the intermediate role of Wikipedia contributors. Consequently, references reflect the interest of a small number of researchers and amateurs who are enthusiastic Wikipedia editors, rather than the general public. Subsequently, Kousha and Thelwall (2017) showed that only 5% of the articles indexed by Scopus between 2005 and 2012 had been cited in Wikipedia, although this percentage rose to 8% when reviews were considered. In contrast, 33% of the academic monographs indexed by Scopus had attracted at least one Wikipedia citation. They concluded that Wikipedia citations were not common enough to be used for impact assessment of articles in most fields. More recently, Teplitskiy et al. (2017) analysed whether journals' impact factor and open access (OA) availability were related to their presence in Wikipedia. They found that a journal's impact factor predicts its appearance in Wikipedia, and that its accessibility increases the odds of being referenced in Wikipedia, although to a lesser extent.

The aim of the current study was to explore the coverage of Library and Information Science (LIS) literature published between 2001 and 2010 in Wikipedia by 2017. The research paid special attention to the methodological issues involved in the use of Wikipedia citations for research evaluation. Specifically, the study aimed:

- to identify the methodological limitations of counting Wikipedia citations,
- to quantify the proportion of LIS literature cited in Wikipedia,
- to analyse the characteristics of Wikipedia entries that cite LIS literature, and
- to measure the OA availability of the LIS articles cited.

Methods

In order to conduct the study, we retrieved the 26,542 articles and reviews indexed in the category "Information Science and Library Sciences" of the *Social Sciences Citation Index* in the *Web of Science* published between 2001 and 2010.

Afterwards, we searched for each of these articles in Wikipedia, and retrieved all the entries in which they were cited. In order to achieve this, we used the advanced search feature of Google, searching for all the words in the article title as an exact phrase and

[3] https://www.altmetric.com/blog/new-source-alert-wikipedia/.

[4] https://help.altmetric.com/support/solutions/articles/6000060980-how-does-altmetric-track-mentions-on-wikipedia.

narrowing the results to those in the domain "wikipedia.org". In the case of articles with very short titles (three or four words), the name of the first author was added to the query and the results were checked manually. All the searches were conducted between the second half of 2016 and early 2017, to allow for an extended period of at least five years from publication of an article. Citation analysis studies usually employ a shorter citation window (impact factors, for instance, are based on the citations received by articles published in the previous two years). However, since this study focuses on citations outside the academic community, an extended citation window seemed appropriate.

Any citation to an article in a Wikipedia entry was recorded either in the "references" section, as "additional reading" or embedded in the text (for instance, in a section of a Wikipedia entry entitled "Example studies that have leveraged the IS success model"). If articles were cited in several Wikipedia entries, all the instances were recorded. Similarly, the citation of an article in different language versions of a single Wikipedia entry was recorded. It should be borne in mind that the different language versions of a single Wikipedia entry are not translations, but are edited independently and therefore may cite different sources.

Finally, when the reference included a link to an external source, we visited the website to find out whether the full-text of the article was available in OA. Again, in the case of articles cited in several Wikipedia entries, all the references were checked, since they may link to different sources. However, in the case of citations to a single article in different language versions of the same Wikipedia entry, only the reference in the English version (or the first retrieved version if the article was not cited in the English version) was checked.

In sum, the research process proceeded as follows:

1. Retrieval of LIS articles and reviews published between 2001 and 2010 indexed in WoS: 26,542 records.
2. Google advanced domain search of LIS articles cited in Wikipedia: 982 citations of 766 articles. For each citation we measured:
    - 2.1 Completeness of the reference: author, title, journal title, DOI, etc.
    - 2.2 Type of Wikipedia entry citing the article.
    - 2.3 OA availability of the cited article when an external link was provided.

Results

Limitations of counting Wikipedia citations

The retrieval of Wikipedia citations to the academic articles in the sample proved to be a difficult task, due to the lack of standardization of bibliographic references. Table 1 provides examples of the most frequently observed problems. Reference 1 illustrates the case of a reference that only includes the article's title and a link to the full-text stored on the publisher's website. Reference 2 shows a slightly more complete citation including the journal and year of publication, in addition to the article title and publisher URL. Meanwhile, Reference 3 includes the author's name, year of publication, title, and URL. In this case, the link leads to a post-print copy of the article deposited in an institutional repository.

Table 1 Examples of incomplete references in Wikipedia

1. Reference including title and URL

• Interpolation of the extended Boolean retrieval model
Source: https://en.wikipedia.org/wiki/Extended_Boolean_model

2. Reference including title, URL, journal and year

• "But the data is already public": on the ethics of research in Facebook, *Ethics and Information Technology*, 2010
Source: https://en.wikipedia.org/wiki/Michael_Zimmer_(academic)

3. Reference including author, year, title and URL

14. ^ Gorman, M. (2001). "Values for Human-to-Human Reference", p179
Source: https://en.wikipedia.org/wiki/Michael_Gorman_(librarian)

4. Reference including authors, year and title

• Walther, Joseph B and D'Addario, Kyle P (2001). The Impacts of Emoticons on Message Interpretation in Computer-mediated Communication
Source: https://en.wikipedia.org/wiki/Social_information_processing_(theory)

5. Reference including authors, year, title, journal, volume, issue, pages and DOI

• Walther, Joseph B and D'Addario, Kyle P (2001). "The Impacts of Emoticons on Message Interpretation in Computer-mediated Communication". Social Sciece Computer Review. **19** (3): 324–347. doi:10.1177/089443930101900307.
Source: https://en.wikipedia.org/wiki/Joseph_Walther

6. Reference including abbreviated journal title

^ *a b* Kim, Matthew; Johnson, Kevin (2002). "Personal health records: evalution of functionality and utilty". *JAMIA*. **9** (2): 171–180. doi:10.1197/jamia.M0978. PMC 344574. PMID 11861632.
Source: https://en.wikipedia.org/wiki/Personal_health_record

7. Erroneous reference in which authors' names and surnames are inverted (the authors are Maryam Alavi, George M. Marakas and Youngjin Yoo)

7. ^ 7.0.7.1 Maryam A. & George M. & Youngjin Y.(2002). A Comprative a Study of Distributed Learning Environment on Learning Outcomes. Information Systems Research, Vol. 13, No. 4
Source: https://en.wikipedia.org/wiki/Social_learning_theory (Hebrew version)

The degree of completeness of the references varies from entry to entry, even for a single article. Reference 4 only includes the names of the authors, year of publication and title, whereas Reference 5 provides a much more detailed citation of the same article obtained from a different Wikipedia entry. Abbreviated journal titles can make even relatively complete references difficult to retrieve, as in Reference 6, in which the *Journal of the American Medical Informatics Association* has been abbreviated to *JAMIA*. Some references contained errors, such as that in Reference 7, in which the authors' names and surnames have been inverted.

The use of the "cite journal" template[5] to create citations for scientific papers is inconsistent. It is common to find Wikipedia entries in which "references" have been edited using the recommended template, but citations included in sections such as "further reading", "select bibliography" or "external links" have not. This is the case, for instance, of References 1 and 2 in Table 1. Even when the citation template is used, examples in Table 1 show that many parameters may be missing. The inclusion of a DOI in the reference could be used to automatically extract Wikipedia citations to academic articles. However, for articles published in 2010, the latest year considered in our study, just 61

[5] https://en.wikipedia.org/wiki/Template:Cite_journal.

references out of 115 (53%) included a DOI. Again, we can find examples of a single article cited in several Wikipedia entries with and without a DOI.

Proportion of LIS literature cited in Wikipedia

Overall, just 2.9% (766 articles) of the LIS output published between 2001 and 2010 and indexed in the *Social Sciences Citation Index* had been cited in Wikipedia by the time of data collection. Since some articles had been cited in several Wikipedia entries, the total number of citations retrieved was 982 (Table 2).

Citations in Wikipedia biographies

As could be expected, Wikipedia entries citing LIS literature were related to topics in the field. Frequently, these Wikipedia entries were biographical articles about well-known LIS scholars (such as Marcia J. Bates, Hope A. Olson and Tom Wilson, to name but a few) describing their education, work and awards, among other information.

Most of these biographical entries include a list of publications authored by the scholar in question (see, for instance, Reference 5 in Table 1). In fact, a total of 13.5% of the Wikipedia citations retrieved in our study were made in biographical entries about one of the authors of the cited article (Table 2). The number of citations in authors' biographical entries was especially significant for articles published in the initial five years covered in our study, although it decreased for more recent literature.

Open access availability of articles cited in Wikipedia

Scholarly journals often require expensive subscriptions. It is therefore questionable whether Wikipedia contributors have access to these sources or whether they rely on OA sources to edit entries. Our results show that 31.2% of the Wikipedia citations were linked to an OA source, with this percentage increasing for more recent literature (Table 2).

At this point, we counted separately the citations to a single article in several Wikipedia entries, since a reference may link to an OA source in one entry, but not in another. For instance, Table 3 shows four different linking options for a single article in four Wikipedia entries: Reference 1 does not include a link; Reference 2 is linked to the publisher's website that requires a subscription to gain access to the full-text; Reference 3 includes a broken link to the co-author's personal website; and Reference 4 links to a freely available post-print copy of the article stored in the Internet Archive version of the page linked in Reference 3.

The 306 references that included an OA link pointed to three kinds of sources in a balanced manner: publishers' websites (fully OA journals; articles that were OA after an embargo period and OA articles in hybrid journals, among others): 39.2%; repositories (disciplinary or institutional): 30.4%; and websites (personal, departmental and social networks, among others): 30.4%.

In the case of citations to articles published in fully OA journals, we could expect to systematically find links to the full-text available on the publisher's website. However, this was not always the case. Table 4 shows two examples of references to articles available in OA journals that were not linked from Wikipedia. The first example corresponds to an article published in *College and Research Libraries*, a journal that is currently available in OA. The reference provides a (broken) link to the social network Academia.edu. A

Table 2 LIS literature cited in Wikipedia by publication year

Year	Articles in WoS	Articles cited in Wikipedia	Citations of LIS literature	Wikipedia citations in authors' biographies	OA citations
2001	2349	51 (2.2%)	77	17 (22.1%)	14 (18.2%)
2002	2337	59 (2.5%)	66	13 (19.7%)	16 (24.2%)
2003	2315	62 (2.7%)	73	17 (23.3%)	16 (21.9%)
2004	2173	80 (3.7%)	111	18 (16.2%)	36 (32.4%)
2005	2499	84 (3.4%)	105	21 (20.0%)	27 (25.7%)
2006	2512	79 (3.1%)	99	12 (12.1%)	30 (30.3%)
2007	2820	88 (3.1%)	122	6 (4.9%)	45 (36.9%)
2008	2962	94 (3.2%)	115	13 (11.3%)	37 (32.2%)
2009	3165	86 (2.7%)	99	9 (9.1%)	36 (36.4%)
2010	3272	83 (2.4%)	115	7 (6.1%)	49 (42.6%)
Total	26,542	766 (2.9%)	982	133 (13.5%)	306 (31.2%)

Table 3 Examples of different linking options to a single article

1. Reference without a link

9. ^ Sondergaard T. F.; Andersen J.; Hjorland B. Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. Journal of Documentation 2003, 59, (3), 278–320

Source: https://en.wikipedia.org/wiki/Grey_literature

2. Reference with a link to a publisher

2. ^ Søndergaard, T. F.; Andersen, J.; Hjørland, B. (2003). "Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model". *Journal of Documentation.* **59** (3): 278 doi:10.1108/00220410310472509

Source: https://en.wikipedia.org/wiki/UNISIST_model

3. Reference with a broken link to the co-author's personal website

Fjordback Søndergaard, T.; Andersen, J. & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. *Journal of Documentation*, *59*(3), s. 278–320. http://www.db.dk/bh/UNISIST.pdf

Source: https://en.wikipedia.org/wiki/Source_literature

4. Reference with a link to a copy of the co-author's personal website in the Internet Archive

- Fjordback Søndergaard, Trine; Andersen, Jack & Hjørland, Birger (2003). Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model. Journal of Documentation, 59(3), 278–320. (Available at: http://web.archive.org/web/20050320083023/http://www.db.dk/bh/UNISIST.pdf)

Source: https://en.wikipedia.org/wiki/Scientific_communication

possible reason for this situation is that the reference, according to the retrieved date, was introduced in December 2010, but *College and Research Libraries* did not become OA until the following year. The second example shows a reference to an article published in *Information Research*, a fully OA journal since its creation. However, the reference does not include a link. Obviously, if publishers' freely available versions are not always linked by Wikipedia contributors, it is highly plausible that copies deposited in repositories or

Table 4 Examples of references to OA articles in Wikipedia without links to the full-text

1. Reference to an article in *College and Research Libraries*
2. ^ Beall, Jeffrey; Kafadar, Karen (2005). "The Proportion of NUC Pre-56 Titles Represented in OCLC Worldcat". College & Research Libraries. **66** (5): 431–5. Retrived 2010-12-23.
Source: https://en.wikipedia.org/wiki/National_Union_Catalog
2. Reference to an article in *Information Research*
5. ^ Shiyong Lu, Dapeng Liu, Farshad Fotouhi, Ming Dong, Robert Reynolds, Anthony Aristar, Martha Ratliff, Geoff Nathan, Joseph Tan, and Ronald Powell, " Language Engineering for the Sematic Web : a Digital Library for Endangered Languages", Information Research, 9(3), April 2004.
Source: https://en.wikipedia.org/wiki/Language_planning

other sources are not linked either, making the results in Table 2 an underestimate of the OA availability of cited sources.

Discussion and conclusions

Among other altmetric indicators, citations in Wikipedia have been proposed as an alternative to traditional impact metrics. Citations of articles in Wikipedia can be seen as a metric that partially captures the societal and educational impact of an article in a wider audience beyond the academic community. However, the results of this study reveal severe limitations in the use of Wikipedia citations for research evaluation purposes.

The lack of standardization of Wikipedia references makes it difficult to measure them with a minimum level of precision. Unlike bibliographies in academic publications, where references are edited to ensure that they are correct, Wikipedia citations are frequently incomplete or even erroneous. Essential fields for the proper identification of articles such as authors' names or journal titles may be missing, making it impossible to retrieve citations. This feature, combined with the absence of document identifiers such as DOIs means that we cannot rely on automatic extractions of citations. If professionally edited citation indexes, such as Scopus and Web of Science, have been criticized for inaccuracies that make it difficult to retrieve some documents and distort bibliometric indicators (Franceschini et al. 2015, 2016), it is hard to consider using Wikipedia citations for research evaluation purposes. Bibliometric indicators based on Wikipedia citations will be unlikely to reach the requirements of robustness and replicability necessary to be used in decision-making processes.

The number of Wikipedia citations is also too small to be used in research evaluation. Less than 3% of LIS articles published between 2001 and 2010 had been cited in Wikipedia by 2016. This figure results from a detailed search of individual articles including manual checks but, given the lack of standardization and incompleteness of many references, any automatic attempt to retrieve Wikipedia citations would probably lead to a lower figure. Given the scarce amount of information provided in some references, it is also possible that we have missed some citations. Although our study focuses on a small discipline such as LIS, the results are consistent with those obtained by Lin and Fenner (2014) who found that just 4% of PLOS articles had been cited in Wikipedia, and Kousha and Thelwall (2017) who reported that only 5% of the articles indexed by Scopus between 2005 and 2012 had been cited in the encyclopedia. The share of LIS articles cited by year of publication remained fairly stable throughout the decade analysed. This issue requires

further study with a larger sample, since Wikipedia citations can be expected to behave differently from those in academic journals or other scholarly outputs. The fact that academic journals are addressed to a scholarly audience while Wikipedia is aimed at the general public may result in a lower level of obsolescence of citations in Wikipedia compared to those in academic articles which tend to cite cutting-edge research.

In addition to the low percentage of scholarly literature articles cited in Wikipedia, attention must also be paid to the representativeness of these citations. As stated by Thelwall (2016), the use of Wikipedia citations as a proxy for public interest in research articles is limited, due to the intermediate role of Wikipedia contributors, with references reflecting the interest of a small number of researchers and amateurs who are enthusiastic Wikipedia editors, rather than those of the general public. Although our study does not deal with this issue, our results reveal some aspects that should also be considered when Wikipedia citation data is interpreted. This is the case of the relatively large amount of Wikipedia citations retrieved in the biographies of articles' authors. Wikipedia biographies of relevant scholars often list their publications, which increases the number of citations received by well-known scholars in the field. This results in a phenomenon of accumulated advantage similar to the Matthew effect. Our results show that this phenomenon is more evident for older literature, which suggests that biographical Wikipedia entries are created for more senior scientists.

The relationship between OA availability and Wikipedia citations is also of interest, since we can intuitively assume that easy accessibility makes articles more likely to be referenced (Teplitskiy et al. 2017). Our results show that 31.2% of the Wikipedia citations of LIS literature linked to an OA source, with this percentage increasing for more recent literature. However, this is probably an underestimate of OA availability due to the incompleteness of Wikipedia citations, and the fact that links to OA sources are frequently missing.

References

Evans, P., & Krauthammer, M. (2011). Exploring the use of social media to measure journal article impact. *AMIA Annual Symposium Proceedings*, 374–381. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243242/. Accessed 2 May 2017.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics, 102*(3), 2181–2186. doi:10.1007/s11192-014-1503-4.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus. *Journal of Informetrics, 10*(1), 174–182. doi:10.1016/j.joi.2015.11.006.

Halfaker, A., & Taraborelli, D. (2015). Scholarly article citations in Wikipedia. *Figshare*. doi:10.6084/m9.figshare.1299540.v8.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics, 101*(2), 1145–1163. doi:10.1007/s11192-013-1221-3.

Huvila, I. (2010). Where does the information come from? Information source use patterns in Wikipedia. *Information Research*, *15*(3). http://www.informationr.net/ir/15-3/paper433.html. Accessed 2 May 2017.

Kousha, K., & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology, 68*(3), 762–779. doi:10.1002/asi.23694.

Lin, J., & Fenner, M. (2014). An analysis of Wikipedia references across PLOS publications. *Figshare*. doi:10.6084/m9.figshare.1048991.v3.

Luyt, B., & Tan, D. (2010). Improving Wikipedia's credibility: References and citations in a sample of history articles. *Journal of the Association for Information Science and Technology, 61*(4), 715–722. doi:10.1002/asi.21304.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. A., & Lanamäki, A. (2015). The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology, 66*(2), 219–245. doi:10.1002/asi.23172.

Nielsen, F. (2007). Scientific citations in Wikipedia. *First Monday*, *12*(8). http://firstmonday.org/article/view/1997/1872. Accessed 2 May 2017.

Stankus, T., & Spiegel, S. E. (2010). Wikipedia, Scholarpedia, and references to journals in the brain and behavioral sciences: A comparison of cited sources and recommended readings in matching free online encyclopedia entries. *Science & Technology Libraries, 29*(3), 258–265. doi:10.1080/0194262X.2010.497711.

Tarango, J., Ascencio-Baca, G., Romo-González, J. R., & Gutiérrez-Balderrama, J. P. (2017). Tendencias de información en la categoría de artículos destacados de Wikipedia: una perspectiva sobre la obsolescencia de los contenidos virtuales y de libre acceso. *Anales de Documentación*. doi:10.6018/analesdoc.20.1.259721.

Teplitskiy, M., Lu, G., & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology, in press.*. doi:10.1002/asi.23687.

Thelwall, M. (2016). Does astronomy research become too dated for the public? Wikipedia citations to astronomy and astrophysics journal articles 1996–2014. *El Profesional de la Información, 25*(6), 893–900. doi:10.3145/epi.2016.nov.06.