Understanding the value of social media metrics for research evaluation

Zahedi, Z.

Author: Zahedi, Z.
Title: Understanding the value of social media metrics for research evaluation
Issue Date: 2018-12-12

Understanding the Value of Social Media Metrics for Research Evaluation

Zohreh Zahedi

Understanding the Value of Social Media Metrics for Research Evaluation

Keywords: Social Media Metrics (Altmetrics); Scholarly Communication; Research Evaluation

Cover design and layout: Andrea Reyes Elizondo
Printed by: Gildeprint-The Netherlands
ISBN: 978-94-6323-401-6

The research described in this thesis was performed at the Centre for Science and Technology Studies (CWTS), Faculty of Social and Behavioural Sciences at Leiden University, in the Netherlands. Part of the research was supported by fundings from the Iranian Ministry of Science, Research, and Technology (MSRT # 89100156), Thomson Reuters (Clarivate Analytics) & Altmetric.com (#1st altmetric grant); and the Leiden University Fund (LUF #4509).

Understanding the Value of Social Media Metrics for Research Evaluation

Proofschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 12 december 2018
klokke 11:15 uur

door

Zohreh Zahedi

geboren te Lar, Iran
in 1980

Promotores en promotiecommissie

Promotor: Prof. dr. Paul Wouters
Co-promotor: Dr. Rodrigo Costas

Leden promotiecommissie:

Prof. dr. Sarah de Rijcke (Leiden University, the Netherlands)
Dr. Thed van Leeuwen (Leiden University, the Netherlands)
Prof. dr. Mike Thelwall (University of Wolverhampton, UK)
Prof. dr. Isabella Peters (University of Kiel, Germany)
Dr. Cassidy R. Sugimoto (Indiana University Bloomington, US)

To all scholars and centers of academic knowledge

Abstract

The availability of indicators based on social media has opened the possibility to track the online interactions between social media users and scholarly entities. Indicators derived from these online interactions reflect aspects such as *how often*, by *whom*, and *when* are scholarly publications mentioned and discussed on social media platforms. These new indicators, popularly known as *altmetrics* and more technically referred as *social media metrics* are usually proposed as potential alternatives to citation-based indicators to inform research evaluation. The research presented in this book provides the state of the art in the possibilities of social media metrics for informing research evaluation. The main ambition is to increase the knowledge and understanding of the limitations, challenges, and actual possibilities of social media metrics for research evaluation. This thesis describes the presence and distribution of different social media metrics across scientific publications and their relationship with traditional impact indicators. It further describes the main characteristics of publications mentioned in Mendeley as one of the main social media metrics platforms. Moreover, critical challenges regarding data quality issues of social media data are thoroughly described and discussed. Finally, some possibilities and applications of social media metrics for informing research evaluation are presented. The research presented in this book provides both empirical and conceptual answers for the consideration of social media metrics in research evaluation.

Contents

Chapter 1 **1**
General Introduction

Chapter 2 **25**
How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications

Chapter 3 **51**
General discussion of data quality challenges in social media metrics: extensive comparison of four major altmetric data aggregators

Chapter 4 **93**
On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications

Chapter 5 **115**
Mendeley Readership as a Filtering Tool to Identify Highly Cited Publications

Chapter 6 **137**
Exploring topics of interest of Mendeley users using VOSviewer overlay visualization approach

Chapter 7 **157**
Discussion, Conclusions, and Further Research

Appendices **179**

Summary 181
Nederlandse Samenvatting 185
Supplementary information to Chapters 3, 4, 5, 6 189
List of publications 215
Acknowledgment 219
Curriculum Vietae 221

CHAPTER 1

General Introduction

Introduction

Developing robust indicators for assessing the impact and value of research has been highlighted as a crucial step to support the process of decision making in the context of research evaluation (Wilsdon et al., 2015; Wilsdon & Al., 2017). Limitations of current citation-based indicators in reflecting the broad value of research (beyond scientific impact) and its contributions to society are seen as critical challenges in research evaluation (MacRoberts & MacRoberts, 2017; Haustein & Larivière, 2015; Moed, 2005; MacRoberts & MacRoberts, 1989; Martin & Irvin, 1983). These limitations of citation-based indicators have led to the development of alternative indicators in research evaluation that could represent broader impacts of research. For instance, some national assessment exercises such as the Research Excellence Framework(REF)[1] in the UK or the Standard Evaluation Protocol (SEP)[2] in the Netherlands have incorporated criteria for evaluating the scientific, social, economic, and cultural impact of research in their respective assessments. The availability of indicators based on social media has opened the possibility to track and measure different aspects of online interactions in social media platforms (including aspects such as how often, by whom, and when scholarly publications are mentioned and discussed on social media). These new indicators, popularly known as *altmetrics* and more technically referred as *social media metrics* (Haustein, Bowman, & Costas, 2016; Wouters, Zahedi, & Costas, 2018) are usually proposed as potential alternatives to citation-based indicators to inform research evaluation (Priem, Taraborelli, Groth, & Neylon, 2010a). However, it is not yet clear how these new indicators can be used for the evaluation of scientific activities, and their validity for such purpose is still a bone of contention in the literature (Sugimoto, Work, Larivière, & Haustein, 2017). The main ambition of this PhD thesis is to increase our knowledge and understanding of the limitations, challenges, and actual possibilities of social media metrics for research evaluation. This chapter presents an introduction to social media and scholarly communication in section 1.1. Section 1.2 discusses the origins, definitions, and data availability (through different altmetric data aggregators) of social media metrics. Section 1.3 reviews the challenges, limitations, and possibilities of social media metrics for research evaluation; section 1.4 describes different social media metrics data sources, particularly focusing on Mendeley as a specific relevant data source for research evaluation. Finally, sections 1.5 and 1.6 introduce the aim and research questions of this thesis.

1.1. Social media and scholarly communication

Social media emerged in recent years as a novel and innovative form of communication among organizations and individuals, including scholars (McCay-Peet & Quan-Haase, 2017; Gruzd, Staves, & Wilk, 2012). Although there is no consensus about the definition of social media in the literature, different definitions share some elements such as *user-generated contents* that are *shared across web-based environments*. In general, the term *social media* refers to “web-

[1] http://www.ref.ac.uk
[2] https://www.knaw.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021

based Internet applications that allow the creation, access, and exchange of user generated content that is ubiquitously accessible" (Batrinca & Treleaven, 2015). Another definition considers them as "web-based services that allow individuals, communities, and organizations to collaborate, connect, interact, and build community by enabling them to create, co-create, modify, share, and engage with user-generated content that is easy accessible" (McCay-Peet & Quan-Haase, 2017, page 16).

The rise of social media is changing how individuals are communicating and exchanging information in modern societies. Uptake of social media by American adults has grown from 5% in 2005 to 69% in 2011 (Pew research[3], 2017). It is estimated that global social media users will be around 3 billion by the year 2021 (Statista[4], 2018). The use of social media is also gaining momentum among academics (van Noorden, 2014) who use these platforms both for personal and professional purposes (Bowman, 2015). In addition to formal scholarly communication channels (i.e., journal publications, books, etc.), academics are also using social media platforms to communicate, collaborate, and disseminate research among various audiences (Gruzd, Staves, & Wilk, 2012). General social media platforms like Facebook, Twitter, Wikipedia, blogs, social bookmarking tools (e.g., Mendeley, Zotero, CiteULike, BibSonomy) (Gruzd & Goertzen, 2013) and particularly, academic social networking sites such as ResearchGate (with 14 million members[5]) and Academia.edu (with 59 million academics sharing over 20 million publications[6]) are becoming popular among researchers and academics (Thelwall & Kousha, 2014, 2015a; Wouters, Zahedi, & Costas, 2018; Wouters et al., 2018). Sharing of knowledge and publications, keeping up-to-date with other researchers, connecting with peers, receiving help to answer questions, following research topics, and being aware of job opportunities are among the motivations for the use of academic social networking sites such as ResearchGate (Chakraborty, 2012).

However, the popularity of social media varies across groups of researchers, and the extent to which these platforms are used for scholarly or personal purposes has been subject of several studies (Sugimoto, et al., 2017). In a survey published in Nature (Van Noorden, 2014) it was shown that Twitter is mostly used for academic purposes (e.g., to comment or to follow discussions about research). Mendeley is used for discovering papers and organizing references while Academia.edu or ResearchGate are mostly used by researchers for maintaining online presence (van Noorden, 2014). The existence of biases in the adoption and use of social media across genders, generations, or countries is also discussed in the literature (Alperin, 2013; Bolton, et al., 2013; Nicholas, et al., 2015; Thelwall & Kousha, 2015a). There are important differences in the perceptions among scholars about the value of social media for scholarly purposes and not all scholars consider social media as trustworthy or creditable channels for formal scholarly communication (Jamali, et al., 2014). For instance, Hank, Sugimoto, Tsou, & Pomerantz (2014) found that faculty members didn't consider Facebook as

[3] http://www.pewinternet.org/fact-sheet/social-media/
[4] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/
[5] https://www.researchgate.net/about
[6] https://www.academia.edu/about

a professional channel to interact with their students. European scholars considered disseminating research via social media platforms as less important than disseminating results via journals/books and conference publications (Jamali, Nicholas, & Herman, 2016). Also, a majority of research policy makers and researchers in higher education in Finland still expressed doubts about how indicators based on social media platforms could be used for research funding applications (Fraumann, 2017).

In spite of these biases and unclear value of social media for research evaluation, social media platforms are increasingly integrated in different phases of the research workflow, from identifying research opportunities to disseminating research across different communities of researchers from all disciplines (Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011). Academic papers are saved in some social reference manager tools such as Mendeley, or mentioned or discussed on Twitter or blogging platforms (Thelwall, Haustein, Larivière, & Sugimoto, 2013). Also, a majority of users of social media platforms such as tweeters (Andrew Tsou, Bowman, Sugimoto, Larivière, & Sugimoto, 2016), bloggers (Shema, Bar-Ilan, & Thelwall, 2012), or Mendeley users (Haustein & Larivière, 2014) are academics or employed in academia (Puschmann & Mahrt, 2012). The study of the activities of scholars on social media has led to the development of tools to identify scientists on social media platforms like Twitter (Costas, van Honk, & Franssen, 2017; Hadgu & Jäschke, 2014; Ke, Ahn, & Sugimoto, 2016). Moreover, it seems that social media strategies[7] are becoming part of the communication policies of academic organizations. Social media channels are increasingly embedded in academic organizations' web pages (libraries, universities, etc.) as well as in publishers and journals websites[8] (Sugimoto et al., 2017). This increasing presence of social media in scholarly activities has resulted in the proliferation of social media data and indicators that could be used to track and measure these activities.

Social media metrics refer to the metrics (e.g., number of tweets, Facebook counts, Wikipedia mentions, blog posts, news mentions, readers in social reference management tools, etc.) for *scholarly objects* (including all kinds of *research products* – (Piwowar, 2013)) driven *from social media platforms* (originally these indicators have been popularly known as *altmetrics* (Priem, 2010; Haustein, Bowman, & Costas, 2016). More recently it was suggested that tracking all these online events opens the possibility of studying the interactions between social media and all sorts of scholarly entities (i.e., not only scholarly objects, but also scholarly actors like authors, universities or journals) (Haustein, et al., 2016; Wouters, Zahedi, & Costas, 2018). Studying the interactions between social media and scientific entities provides the opportunity for obtaining valuable insights into the communications between social media audiences and scholarly actors. For instance, nowadays it is possible to study how scientific documents are discussed online or how different users (academic or not) engage with scholarly content and share it with different audiences. The study of "*the relationships and interactions between science and social media*" as sources of information instead of just

[7] https://university-relations.umn.edu/resources/social-media-strategy-and-best-practices; https://www.elsevier.com/authors-update/story/social-media/making-the-most-of-social-media
[8] http://journals.plos.org/plosone/

impact indicators has led to the proposal of the *social media studies of science* as a new approach to obtain a more comprehensive perspective of the study of scholarly communication in the age of social media (Costas, 2017; Wouters, et al., 2018).

1.2. Scholarly indicators based on social media: origins and developments

As explained in the previous section, social media metrics (popularly called *altmetrics*) refer to the metrics (number of tweets, Facebook counts, Wikipedia mentions, blog posts, news mentions, readers in social reference management tools) for scholarly objects that are obtained from social media platforms. In this section the origin of the concept of *altmetrics*, the development of *altmetric data aggregators*, and the main conceptual debates around altmetrics are presented and discussed.

The origin of altmetrics, its promises, and drivers

The concept of *altmetrics* (Priem, 2010) originated from the desire to improve and complement the traditional impact assessment tools of research evaluation. The '*Altmetrics manifesto'* based itself on the diverse users' engagements with scholarly contents in the social web and across different online platforms (Priem, Taraborelli, Groth, & Neylon, 2010). The expectation of the *Altmetrics manifesto* was to enable real-time monitors for impact assessment (Priem & Hemminger, 2010) and to enhance the traditional peer review by enabling "*soft peer review*" (Taraborelli, 2008). Two important elements are central in the Altmetrics manifesto. First, the idea of enlarging the set of research products credited for research evaluation, by including outputs such as blog posts, datasets, codes, etc. (Piwowar, 2013). Second, the possibility of measuring more diverse forms of real-time impact across diverse audiences (researchers, general public, clinicians, practitioners) able to complement traditional scholarly impact analyses (Priem, Piwowar, & Hemminger, 2012).

One of the driving forces in the development of altmetrics was the 'promise' of these new metrics to solve the inadequacies of the more traditional and established scholarly metrics (particularly citations and peer review) (Wouters & Costas, 2012). This was in line with the need for more multi-dimensional research performance evaluations that was emphasized in the literature (Rousseau & Ye, 2013; Cronin, 2014; Van Leeuwen, Visser, Moed, Nederhof, & Van Raan, 2003). This need for more multi-dimensional research evaluations is grounded in the 'gap' existing between the actual value or quality of the scientific work and the way it is assessed in research evaluation. Wouters (2014)[9] introduced the notion of the *evaluation gap* as the "gap between […] the dominant criteria in scientific quality control […] and on the other hand the goals of the research under evaluation or the roles of research in society". In the same line Cronin (2014) also argued that "a scholar's work may well have a range of impacts over time in different contexts, with different audiences, and for different reasons and traditional bibliometric indicators may not fully reflect these multivariate contributions"

[9] https://citationculture.wordpress.com/2014/08/28/a-key-challenge-the-evaluation-gap/

(Cronin, 2014, page. 14). Thus, the new altmetric indicators could be seen as potential alternatives that could fill this gap.

A more technical driving force for the growth of altmetrics has been the development of Information and Communication Technologies (ICT), particularly the development of advanced Application Programming Interfaces (APIs)[10] which has increased the capacity of collecting and analyzing large sets of social media data by altmetric data aggregators and researchers. In addition, the open science movement contributed to the development of altmetrics (Moed, 2017) by demanding an increased transparency and availability of research to the general public (Friesike & Schildhauer, 2015). Open science has been defined as "*the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process"* (Nielsen, 2011). Open science is also concerned with the idea that access to any research products should be open and freely available (Holmberg, 2017). Wide access to research products is supposed to increase their visibility and hence facilitate the dissemination, sharing, and publication of data and research in online environments.

Finally, in recent years altmetrics have also received a lot of attention from research funders and science policy makers, due to their potential to demonstrate the contribution that scientific research has had to society-at-large (Wilsdon, et al., 2015). However, there is not yet enough evidence in the literature to show how altmetrics can actually reveal the societal impact of research (Tsou, Bowman, Ghazinejad, & Sugimoto, 2015; Alperin & Reilly, 2017). Further studies are needed to explore the potential and importance of these new social media metrics for research impact assessments (Wouters & Costas, 2012). Exploring these potentials are the main focus of this PhD thesis.

The rise of altmetric data aggregators

A critical element in the development of altmetrics and altmetric indicators can be attributed to the increasing availability of altmetric data and indicators, particularly after the foundation of several *altmetric data aggregators* (see Table 1). One of the first altmetric data aggregators was the *Lagotto open source Article-Level Metrics application*- formerly called *PLOS Article-Level-metrics (ALM)*. It started in 2009 by providing the mentions in social media (Twitter, Facebook, etc.) of PloS articles, and later on for any article from any publisher (Fenner, 2013). *Impact Story* (formerly called 'Total Impact') began in 2011 as an open source tool that provided aggregated impact of diverse research products across several social media tools. Other altmetric data aggregators include *Altmetric.com* (Adie & Roe, 2013), *Plum Analytics* (Buschman & Michalek, 2013), and more recently *CrossRef Event Data* (Wass, 2017). All of them provide aggregated metrics for scholarly materials coming from different social media sources. Table 1 provides an overview of the social media sources covered and indicators provided by these aggregators. Moreover, there are other online platforms such as

[10] Application Programming Interface (API) is "a set of subroutine definitions, protocols, and tools for building application software. In general terms, it is a set of clearly defined methods of communication between various software components". https://en.wikipedia.org/wiki/Application_programming_interface

Academia.edu, ResearchGate.net, Microsoft Academic Scholar Universe.com, SemanticScholar.org, or Loop (https://loop.frontiersin.org/) that also provide indicators from social media sources or with a social media component for researchers and their outputs (Orduña-Malea, Martín-Martín, & López-Cózar, 2016; Thelwall & Kousha, 2017; Costas & Franssen, 2018).

The existence of these different altmetric data aggregators and the proliferation of indicators provided by them at different aggregation levels (outputs, individuals, institutes, countries, etc.) raise some important questions such as *from where*, *when*, and *how* social media metrics are collected, processed, and reported by these aggregators. Hence, it is critical to explore to what extent altmetric aggregators differ in the social media metrics provided and to understand the reasons for (dis)similarities across them, and how they can influence the conceptual and analytical possibilities of the metrics provided.

Table 1. Overview of altmetric aggregators, data sources, and metrics.

Aggregators	Founded year	Founded by	Category of impact	Metrics, raw data and data sources
Lagotto (PLOS ALM)	2009	Public Library of Science	Views, saves, citations, recommendations, discussions	Views from PMC usage; saves in Mendeley and CiteULike; discussions in Twitter, Facebook, Nature blogs, ScienceSeeker, ResearchBlogging, Wordpress.com, Wikipedia, Reddit, and OpenEdition; citations from CrossRef, PubMed, PMC, and DataCite; recommendations from F1000.
Impact Story	2011	Jason Priem and Heather Piwowar	Buzz and engagement (volume and quality of discussion around artifacts), openness	Altmetric.com, Mendeley, Twitter, CrossRef, ORCID, Base.
Altmetric.com	2011	Digital Science company	Altmetric Attention Score, mentions and readers	Twitter, Facebook, policy documents, Wikipedia, news, blogs, Mendeley, Pub peer and Publons, Faculty of 1000 Prime Reddit, Stack overflow, Google Plus, YouTube, Open Syllabus Project, Scopus and Web of Science citations.
Plum Analytics	2012	Andrea Michalek, acquired by EBSCO in 2012 and by Elsevier in 2014	Usage, captures, mentions, social media, and citations	Usage category includes abstract and full text views, downloads, URLs clicks, Dryad, figshare, and Slideshare views, Github collaborators, WorldCat holdings, Vimeo, YouTube, SoundCloud plays, Link Outs; Capture category includes Delicious bookmarks, Mendeley, CiteULike, and Goodreads readers, Slideshare, SoundCloud, and YouTube favorites, Github followers, forks, and watchers, Vimeo and YouTube subscribers, exports and saves in EBSCO; Mention category includes (economic) blog mentions, Reddit, Slideshare, Vimeo, and YouTube comments, Forum topic counts in Vimeo, Gist count in Github, news mentions, Wikipedia and StackExchange links, Amazon, Goodreads, and SourceForge reviews; Social media category includes Vimeo and YouTube likes, Google Plus +1, Facebook shares, likes, and c comments, Amazon, Goodreads, and SourceForge ratings, Figshare and SourceForge recommendations, Reddit scores, Twitter; Citation category includes citations from CrossRef, PubMed, RePec, SciELO, SSRN, Scopus, USPTO, clinical and policy citations.

Aggregators	Founded year	Founded by	Category of impact	Metrics, raw data and data sources
CrossRef Event Data	2017	*CrossRef*	Discussed, mentions, annotations, references, links, citations	Discussed in blogs and media, Wordpress.com, Newsfeed, Reddit and links in subreddits, StackExchange sites; Mentions in tweets; Annotations in Hypothes.is; References on Wikipedia pages; Links to DataCite and CrossRef registered contents; Citations in Patents via Cambia Lens.

Altmetrics: terminology and definition

Priem originally defined altmetrics as "*the creation and study of new metrics based on the Social Web for tracking, analyzing, and informing scholarship or as a form of information filtering tool*" (Priem, et al., 2010). NISO defined altmetrics as "*online events derived from activity and engagement between diverse stakeholders and scholarly outputs in the research ecosystem*" (National Information Standards Organization, 2016). Altmetrics have been also defined as "traces of computerization of the research process, and as a tool for the practical realization of the ethos of science and scholarship in a computerized or digital age" by Moed (2015). Haustein, Bowman, & Costas (2016) defined altmetrics as "*events on social and mainstream media platforms related to scholarly content or scholars, which can be easily harvested (i.e., through APIs), and are not the same as the more 'traditional' concept of citations*" as *social media metrics*". These authors also introduced a framework for social media metrics in which the various types of acts that occur between the 'scholarly objects' and 'agents' in online platforms are grouped into three categories: acts of access (viewing, downloading, and saving), appraise (mentioning, rating, discussing, commenting, or reviewing), and apply (using, adapting, or modifying), depending on the degree of engagement of the agent with the scholarly object. Although there is no consensus among altmetrics researchers[11], most definitions revolve around the *interactions* and *engagement* between actors (scholarly or not) with objects (usually scholarly-related) and/or with other actors (scholarly or not). Thus, the definition of *social media metrics* proposed by Haustein et al. (2016) seems to be more convenient (i.e., in terms of accuracy and inclusiveness of most indicators) over the more popular but more vague term of altmetrics. In this thesis the term *social media metrics* is thus preferred over altmetrics, aligning also with the suggestion by Wouters, et al. (2018)[12].

[11] Since its introduction, the term altmetrics has been criticized as a "good idea but a bad term" (Rousseau and Ye, 2013). An important source of criticism is related to the impossibility to come up with a homogenous definition of what altmetrics are. This is particularly important given the multiplicity of sources and metrics that are included under this umbrella term (Haustein, Bowman, & Costas, 2016). Often, are the altmetric data aggregators who determine what constitutes altmetrics. For example, Altmetric.com and Plum Analytics both cover citations from Policy documents as a form of altmetrics, however Policy document citations can be just seen as another form of citations (Wouters, et al., 2018). Hence, a more accurate terminology for altmetrics is demanded in the literature (Haustein, Bowman, & Costas, 2015).

[12] However, in the chapters of this thesis that refer to published work and where the term altmetrics was used, the terminology originally used in the original publication is respected.

1.3. Challenges, limitations, and possibilities of social media metrics for research evaluation

The proliferation of indicators on the various users' interactions with scholarly outputs on social media platforms introduces new conceptual and practical challenges on how to determine their meanings, values, and possibilities for informing research evaluation and research policy. In this section these limitations and challenges are discussed and some of the possibilities that these new indicators can offer are introduced once their limitations are understood and accommodated.

Limitations and challenges

Priem (2014) presented the "*lack of theory, ease of gaming, and possible biases"* as the three main limitations of altmetrics (Priem, 2014). More recently, *heterogeneity* (the diversity of sources and activities and lack of conceptual frameworks for altmetrics), *data quality* (lack of accuracy, consistency, and replicability of altmetric data), and *dependencies* (on object identifiers, like DOIS and the availability of APIs from social media sources) have been highlighted as the *grand challenges* of social media metrics (Haustein, 2016). Building on these previous discussions, in this thesis three major challenges regarding social media metrics are further elaborated: *conceptual challenges*, *technical and data related challenges,* and *use challenges*.

Conceptual challenges are directly related with the lack of theories or frameworks in which social media metrics can be understood and discussed. The main conceptual challenge of social media metrics is to understand what they are actually measuring. Due to their heterogeneity (Haustein, 2016) as they capture many different activities and acts (Haustein, Bowman, & Costas, 2016), it is impossible to develop a common definition for all altmetrics. Therefore, further research is necessary in order to gain a better understanding of the different platforms and related activities associated with altmetrics, and defining what exactly these indicators reflect.

Technical and data-related challenges have to do with the data quality of social media metrics provided by altmetric data aggregators and their dependency on the technical affordances of underlying platforms and on the availability of the document's identifiers. The development and application of social media metrics is dependent on the characteristics and quality of the underlying data. Different altmetric data providers collect, aggregate, or report the same social media metrics differently (Chamberlain, 2013; Jobmann et al., 2014; Zahedi, Fenner, & Costas, 2014). These technical and methodological differences introduce challenges for interpreting the metrics provided, since it is not clear how these technical and methodological choices may affect the meaning of the indicators. Moreover, social media metrics are limited by the technical availabilities of the platforms (for example the API limits imposed by platforms such as Twitter or Facebook) used to collect the data (Haustein, 2016). In comparison with citation indicators, social media metrics are more prone to gaming by users. Most social media platforms have an open and unsupervised nature. The lack of control in the metrics recorded from the activities of their users increase the possibilities of manipulation and gaming by users (e.g., when the same paper is shared, liked, or tweeted several times by the same user, cf. Robinson-Garcia et al, 2017); or the activities created by bots or cyborgs –(Haustein, Bowman, Holmberg et al., 2016)). Furthermore, the reliance on the availability of document identifiers (such as DOIs, PMID, arXiv id's, etc.) and APIs bias social media metrics towards documents with DOIs or other identifiers (e.g., PMID, arXiv id's, etc.), excluding those without such identifiers.

Use challenges are related with the practical uses that social media metrics can have in real life practices, particularly when using them without considering the context. Examples of contextual elements that are important to consider include aspects such as *who are the users?* (e.g., who are the tweeters, the Mendeley readers, etc.), *what are their motivations?* (e.g., do they tweet a paper to praise it, to criticize it, to mock it?), or *how do they engage with the research objects?* (e.g., is the publication retweeted, is there anything added to say something about the publication? etc.) (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Robinson-Garcia, Costas, Isett, Melkers, & Hicks, 2017). For instance, tweets to academic papers may originate from automated accounts or bots (Haustein, Bowman, Holmberg et al., 2016). Also tweeters' motivations for tweeting about science may range from serious discussions, humorous interactions, or mere self-promoting mentions of publications (Haustein, Bowman, Holmberg et al., 2016). It has also been discussed that most tweets about research are actually very superficial or repetitive (Robinson-Garcia et al., 2017), or with a low engagement of the tweeters with the publications, with the majority of tweets just briefly summarizing the title of the papers (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013; Haustein, 2018).

All of the challenges above call into question the value and validity of social media metrics for research evaluation purposes. Actually, most current social media metrics are not yet used in any formal research assessment (Moed, 2017; Wouters et al., 2015; Wouters et al., 2018; European Commission, 2017; Thelwall & Kousha, 2015b). Nevertheless, these limitations and challenges make it necessary to deeply understand them before considering the applications of social media metrics in any context. A proper understanding of the nature and characteristics of these social media metrics, particularly regarding their value for scientific communication, may still open the chance for their careful inclusion in specific new evaluation contexts, as argued in Wouters et al (2018).

1.4. Possibilities of social media metrics

Despite the challenges discussed above, social media play an important role in communication. Their value for communication and interaction with different communities of users should not be ignored.

The possibility of social media metrics for tracking interactions between social media users and scholarly entities opens four main venues of application of social media metrics with possibilities to inform research evaluation:

- *Social media metrics as indicators of presence and reception of research* (Costas, Van Honk, Calero-Medina, & Zahedi, 2017)*,*
- *Social media metrics as indicators of thematic interest, and local or global reach* (Costas, Van Honk, Calero-Medina, & Zahedi, 2017; Haustein, Bowman, & Costas, 2015)*,*
- *Social media metrics as a form of capturing societal impact* (Bornmann, 2012), and
- *Social media metrics as an early predictor of future citation impact* (Thelwall & Nevill, 2018).

These possibilities are discussed in depth in the following sections.

Social media metrics as indicators of presence and reception of research.
Identifying the coverage of publications mentioned on social media platforms and average number of social media mentions per publications are examples of basic indicators that inform the *presence* and reception of research across different social media sources (Costas et al., 2017). These descriptive indicators allow to compare the research produced by different units of analysis (universities, institutions, research groups, or countries). This type of information could inform research managers about the visibility of research of their units across different social media platforms.

S*ocial media metrics as indicators of thematic interest, and local or global reach.*
Social media metrics can help to unveil *communities of attention* around scholarly documents and scientific topics (Haustein et al., 2015), and to track the local or global reach of scholarly documents and topics across different audiences (Costas et al., 2017). The analysis of the thematic orientation and topics of interest of social media users can depict academic and public interests in science as well as their reception across different geographical locations (Wouters et al., 2018; Zahedi & Costas, 2017). These possibilities enable the characterization of social media users and their use of scholarly content, as well as the identification of typologies of users based on their social media behaviour (Haustein et al., 2015). These type of analyses can inform research managers of hot topics and trends discussed in social media platforms as well as specific groups of users that are interacting with their research (Costas, van Honk,Clara-Medina, & Zahedi, 2017; Wang, Fang, Li, & Guo, 2016).

Social media metrics as a form of capturing societal relevance or impact.
The demand for demonstrating the societal relevance of research or contribution of research to society is quite central in several national research evaluation frameworks (e.g., the UK Research Excellence Framework or the Dutch Standard Evaluation Protocol), as well as for

several funding bodies (Wilsdon et al., 2015). There is no consensus in the literature on the definition of societal relevance or societal impact of research (Bornmann, 2012). S*ocietal relevance* is defined by Meijer (2012) as the "*result of analyzing and measuring productive interactions of scientific research with the non-scientific stakeholder which are professionals, public, private, and scientists groups that could bring social, cultural, economic, and scientific returns based on (as a result of) their interactions with research groups*". Based on this definition, a general framework for scientific and societal (socio-economic) relevance has been proposed (Meijer, 2012). This framework puts an emphasis on the knowledge exchange between research and its related professional, public, and economic contexts and how it connects research to societal issues with health and education sectors and lay public. In the context of REF (2014), societal impact of research is defined as "*where the effect or influence [of research] reaches beyond scholarly research, e.g., on education, society, culture or the economy. Research has a societal impact when auditable or recorded influence is achieved upon non-academic organization(s) or actor(s) in a sector outside the university sector itself – for instance, by being used by one or more business corporations, government bodies, civil society organizations, media or specialist/professional media organizations or in public debate*" (Wilsdon et al., 2015).

The relevance of social media metrics for the measurement of the societal impact or social reach of research lies in its promise to reflect broad impact of research beyond scientific impact (Priem, 2010). This is usually related to the idea that *"bibliometric indicators do not provide any insights on the social or economic impact of research and are, thus, limited to assessing the impact of research within the scientific community"* as emphasized by Haustein & Larivière (2015). Aligning with this idea, tracking the use of research by different academic or non-academic users (public, practitioners, professionals, etc.) on social media platforms could reflect the relevance of research for these users. For instance, it is suggested that tweets to papers (Bornmann, 2014) or citations of papers in policy documents[13] as indicators of the value of research for society or for policy making could reflect 'societal impact' of research (Bornmann, Haunschild, & Marx, 2016). However, tweets to papers have not yet provided any evidence of societal reach of research as the papers were discussed mainly by scholars than users from the general public (Tsou, Bowman, Ghazinejad, & Sugimoto, 2015). Alternatively, it has been suggested that some other approaches such as using semantics and natural language analysis can track the spread of scientific ideas in society (Taylor, 2013). Also, mapping the context of social interactions and communication patterns of researchers in Twitter could better reflect the orientation of researchers towards academia or other stakeholders (Ràfols, van Leeuwen, & Robinson-García, 2017) than just considering mentions of papers on Twitter. Nevertheless, due to the lack of agreement of what is exactly considered as societal impact of research and how to exactly measure it, it has been argued that social media metrics have not yet shown any evidence of the societal impact of research (Alperin & Reilly, 2017).

[13] Since policy mention are not obtained from social media platforms, these measures are actually no social media metrics (Haustein, et al., 2016; Wouters, Zahedi, & Costas, 2018)

Social media metrics as an early impact indicator or predictor of future citation impact. Identifying early or predicting future impacts is an important element that could support research evaluation decisions. The delay between publication dates and citations undermines the use of citations for the analysis of recent publications. Exploring whether social media indicators could reflect early or future citation impacts of publications has been subject of some studies (Eysenbach, 2011; Shuai, Pepe, & Bollen, 2012; Thelwall & Nevill, 2018; Thelwall, 2018). The correlations between citations and social media metrics differ by type of metrics. For instance, the recommendations received by publications from the F1000 post-publication peer-review platform have a weak correlation with citations received by these publicaitons (Bornmann & Leydesdorff, 2013). The low coverage of Web of Science publications in F1000 (Waltman & Costas, 2014) made the F1000 recommendations less relevant for prediction of citations. Tweets to papers in online medical journals (Eysenbach, 2011) and arXiv preprint downloads in physics (Brady, Harnad, and Carr, 2006) correlate moderately with later citations. Blogged articles published in PloS journals tend to receive more citations than non-blogged articels (Shema, Bar-Ilan, & Thelwall, 2014). Finally, the number of Mendeley readers has the strongest correlation with citations of all social media metrics (Costas, Zahedi, & Wouters, 2015). Mendeley has been suggested to provide evidence of early impact of publications. Scopus publications in the field of Library and Information Science received more Mendeley readers than citations in the first months of their online publications (Maflahi & Thelwall, 2016). Mendeley has also been discussed as a useful tool to indicate early evidence of scholarly impact (Thelwall, 2017b).

1.5. Social media metrics data sources

Social media metrics data sources vary based on their scholarly orientation (Mendeley readership, F1000 recommendations, or Wikipedia citations), social media orientation (tweets or Facebook counts), or combination of both (RG score from ResearchGate or counts of publications, citations, downloads, views from Academia.edu) (Wouters et al., 2018). Some of these tools provide open APIs (e.g., Mendeley, Facebook, Twitter) to retrieve the metrics, while others (e.g., Research Gate, Academia.edu) do not.

Of these data sources, *Mendeley* (http://www.mendeley.com) deserves special attention. It is a free online reference manager and academic social network tool founded in 2007 and acquired by Elsevier in 2013. Over 6 million users worldwide use this platform[14]. *Readership counts* provided by Mendeley include the total number of users who have saved (added) a document to their private libraries. Besides, Mendeley offers some statistics on the academic status (students, professors, researchers, librarians, professionals, etc.), discipline and country of the users, as well as tags assigned to the saved publications by them[15]. Readership data in Mendeley can be obtained via an open API[16] and hence large scale data collection from this platform is feasible. The existence of both social and scholarly related features (e.g., saving

[14] https://www.mendeley.com/research-network/community
[15] https://www.mendeley.com/reference-management/stats
[16] http://dev.mendeley.com/methods/

papers, highlighting text, writing notes, sharing or recommending papers, joining relevant groups for scholarly discussions and communications, discovering what others are reading, etc.) (Gunn, 2013) makes Mendeley an important source of both scholarly and social media data. Moreover, its higher coverage of publications than any other social media platforms (Haustein et al., 2013; Thelwall, 2017), its popularity among academic and non-academic users (Haustein & Larivière, 2014; Mohammadi, Thelwall, Haustein, & Larivière, 2015; Zahedi, Costas, & Wouters, 2013), and the moderate correlation of Mendeley readership with citations (Sugimoto et al., 2017) present Mendeley as the most relevant and promising source of social media metrics that is worth to further explore. For these reasons, Mendeley readership is specifically studied in several chapters of this thesis.

1.6. Aim of this PhD thesis

The main aim of this PhD thesis is to explore the possibilities of social media metrics to inform research evaluation. As mentioned in the introductory section, this work aims to contribute to the state of the art in social media metrics research by providing novel insights into how scholarly objects are covered across multiple social media platforms, the characteristics of the covered publications, and how social media metrics are related to traditional metrics. How different scientific disciplines are over or underrepresented in social media platforms is also analyzed. Moreover, critical challenges regarding data quality issues of social media data are thoroughly described and discussed. The final purpose of this PhD thesis is to depict the actual possibilities of social media metrics (particularly Mendeley readership metrics as explained in section 1.5) in order to establish their potential uses for informing research evaluation.

1.7. Main research questions and structure of the thesis

The main question that this thesis addresses is *what is the potential usefulness and added value of social media metrics to inform research evaluation?*.

This thesis describes the presence and distribution of different social media metrics across scientific publications and their relationship with traditional impact indicators. It compares main altmetric data aggregators in terms of data quality of the social media metrics data provided by them. This thesis further studies the main characteristics of publications mentioned in Mendeley as one of the main social media metrics platforms. Finally, some possibilities and applications of social media metrics (readership metric) are studied. All the specific research questions presented together provide answers for the main research question of this thesis.

Q1: **What aspect of research impact do social media metrics reflect? In particular, how do social media metrics relate to the more traditional bibliometric indicators?**
The answer to this first question is crucial for understanding what social media metrics are and whether and how these metrics relate to citation impact indicators. Chapter 2 provides a general overview of the presence and coverage of publications presented in social media

platforms and the distribution of social media metrics across fields, publication years, and document types. This chapter gives some important insights into the extent to which scientific publications are presented across social media platforms, the amount of social media attention received by them, and disciplinary differences in their reception of social media metrics. This chapter also describes the relationship between social media metrics and citation indicators. Correlation and factor analysis are used in this chapter in order to study the underlying dimensions of these indicators and their relationship with citation indicators.

Q2: **What are the most important challenges regarding data quality in the social media metrics offered by different altmetric data aggregators?**
The second question of this PhD thesis deals with exploring the data quality of social media metrics provided by different altmetrics aggregators. Understanding the underlying reasons of the existing differences in the metrics across altmetrics aggregators is central for the proper development of applications of social media metrics based on these data. Chapter 3 provides a thorough analysis of the most important data quality challenges and issues regarding social media data provided by the major altmetric data aggregators. This chapter presents how the data collection and reporting approaches of these altmetric data aggregators both technically and conceptually influence the metrics provided. The results of this chapter help in gaining an understanding of how the methodological choices in the tracking, collecting, and reporting of altmetric data influence the reported metrics and their analytical possibilities. Important recommendations for the users of social media metrics and data aggregators are proposed and discussed.

Q3: **What are the main characteristics of publications saved and read on Mendeley?**
This third question (discussed in Chapter 4) expects to understand the relationship of different document characteristics that relationship Mendeley readership (as one of the most prominent social media metric) and how the relationship between these different bibliographic characteristics and readership is (di)similar to that observed for citations. Chapter 4 describes the disciplinary differences in the relationship between Mendeley readership and citation counts with particular documents' bibliographic characteristics across a dataset of 1.3 million publications from the Web of Science. The association between Mendeley readership, citation counts, and document characteristics (i.e., document types, number of pages, length of titles, length of reference lists, number of authors, institutes and countries) has been investigated using Ordinary Least Square (OLS) regression analysis. As the OLS model takes into account very high values which are typical for skewed distributions, it is considered as the most suitable regression strategy for altmetrics data (Thelwall & Wilson, 2014; Thelwall, 2016). The chapter contributes to the identification of document-related differences between Mendeley readership and citations. This information is useful for the future construction of appropriate and meaningful indicators based on Mendeley readership.

Q4: **What are the practical analytical possibilities of Mendeley readership metrics?**
The fourth research question of this thesis deals with understanding analytical possibilities of readership metrics for evaluation perspective. This is particularly crucial for any practical application of this type of indicator. Chapters 5 and 6 discuss analytical possibilities of Mendeley readership metrics by focusing on two main sub-questions:

Q4.1: **Do Mendeley readership metrics offer any advantage over journal-based indicators in identifying highly cited publications?**
Chapter 5 presents the results of a large-scale analysis of the distribution and presence of Mendeley readership scores over time and across disciplines across 9.1 million publications from Web of Science from the years 2004-2013. Using precision-recall analysis the chapter studies whether Mendeley readership scores can identify highly cited publications more effectively than journal citation scores.

Q4.2: **What are the topics of interest of different Mendeley users and how do their use of scholarly documents reflect different types of impact of research?**
Chapter 6 focuses on the different user groups in Mendeley and their thematic orientations. A dataset of 1.1 million Web of Science publications from the year 2012 are analyzed. The disciplinary differences in the reading (saving) patterns of different Mendeley user groups are depicted using VOSviewer maps. Topics of interest of different user groups in Mendeley are analyzed in order to identify the topics focused by different communities of users. The results provide important evidence on the use of scientific publications by user groups (particularly non-publishing ones, such as students or librarians). The results of this chapter support the possibility of using users-based readership to inform different types of impact (scientific, educational, and professional) of scientific publications.

Finally, chapter 7 includes the **discussion and conclusions** of the main results of this PhD thesis. It presents the summary of findings and the implications of the results for using them in research evaluation, together with perspectives for further research.

1.8. References

Adie, E., & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, *26*(1), 11–17.

Alperin, J. P. (2013). Ask not what altmetrics can do for you, but what altmetrics can do for developing countries. *Bulletin of the American Society for Information Science and Technology*, *39*(4), 18–21. http://doi.org/10.1002/bult.2013.1720390407

Alperin, J. P., & Reilly, K. (2017). From opportune indicators to narratives: conceptualizing altmetrics for social impact. In *4: AM conference,* 24-26 Septemeber 2017, Toronto, Canada.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY*, *30*(1), 89–116. http://doi.org/10.1007/s00146-014-0549-4

Bolton, R. N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., … Solnet, D. (2013). Understanding Generation Y and their use of social media: a review and research agenda. *Journal of Service Management*, *24*(3), 245–267. http://doi.org/10.1108/09564231311326987

Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports*, 13(8), 673–676. http://doi.org/10.1038/embor.2012.99

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact : A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, *8*(4), 935–950. http://doi.org/10.1016/j.joi.2014.09.007

Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: how often is climate change research mentioned in policy-related documents? *Scientometrics*, *109*(3), 1477–1495. http://doi.org/10.1007/s11192-016-2115-y

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, *7*(2), 286–291. http://doi.org/10.1016/j.joi.2012.12.003

Bowman, T. D. (2015). Investigating the Use of Affordances and Framing Techniques By Schola Rs To Manage Persona L and Professional Impress Ions on Twitter, (July), 256. Dissertation, Indiana University, Bloomington, IN, USA.

Buschman, M., & Michalek, A. (2013). Are alternative metrics still alternative? *Bulletin of the American Society for Information Science and Technology*, *39*(4), 35–39. http://doi.org/10.1002/bult.2013.1720390411

Chakraborty, N. (2012). Activities and reasons for using social networking sites by research scholars in NEHU: A study on Facebook and ResearchGate. *Planner-2012*, 19–27.

Chamberlain, S. (2013). Consuming Article-Level Metrics: Observations and Lessons, *2*(2).

Costas, R. (2017). Towards the social media studies of science: social media metrics, present and future. *Bibliotecas. Anales de Investigación*, *13*(1), 1–5.

Costas, R., & Franssen, T. (2018). On the cautionary use of the h-index and the proliferation of individual-level bibliometric sources: response to Teixeira da Silva & Dobrénski. *Scientometrics*, 115(2):1125-1130. http:/doi.org/10.1007/s11192-018-268

Costas, R., Van Honk, J., Calero-Medina, C., & Zahedi, Z. (2017). Discussing practical applications for altmetrics: development of “social media profiles” for African, European and North American publications. In *22nd International Conference in Science & Technology Indicators (STI),* 6-8 September, 2017, Paris, France.

Costas, R., van Honk, J., & Franssen, T. (2017). Identifying scholars on Twitter: opening the path to the social media studies of science. In *4 AM Conference*, Toronto, Canada.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. http://doi.org/10.1002/asi.23309

Cronin, B. (2014). Scholars and Scripts, Spoors and Scores. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multidimensional indicators of performance* (pp. 3–21). Cambridge, MA: MIT Press.

Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research*, *13*(4), e123. http://doi.org/10.2196/jmir.2012

Fenner, M. (2013). What Can Article-Level Metrics Do for You? *PLoS Biology*, *11*(10), e1001687. http://doi.org/10.1371/journal.pbio.1001687

Fraumann, G. (2017). *Valuation of altmetrics in research funding*. University of Tampere. Retrieved from http://tampub.uta.fi/handle/10024/101653

Friesike, S., & Schildhauer, T. (2015). Open science: many good resolutions, very few incentives, yet. In I. M. Welpe, J. Wollersheim, S. Ringelhan, & M. Osterloh (Eds.), *Incentives and Performance. Givernance of Research Organizations.* Springer.

Gruzd, A., & Goertzen, M. (2013). Wired academia: Why social science scholars are using social media. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3332–3341. http://doi.org/10.1109/HICSS.2013.614

Gruzd, A., Staves, K., & Wilk, A. (2012). Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model. *Computers in Human Behavior*, *28*(6), 2340–2350. http://doi.org/10.1016/j.chb.2012.07.004

Gunn, W. (2013). Social Signals Reflect Academic Impact: What it Means When a Scholar Adds a Paper to Mendeley. *Information Standards Quarterly*, *25*(2), 33. http://doi.org/10.3789/isqv25no2.2013.06

Hadgu, A. T., & Jäschke, R. (2014). Identifying and analyzing researchers on twitter. *CEUR Workshop Proceedings*, *1226*, 164–165. http://doi.org/10.1145/2615569.2615676

Hank, C., Sugimoto, C. R., Tsou, A., & Pomerantz, J. (2014). Faculty and student interactions via Facebook : Policies , preferences , and practices, *56*(5), 216–223. http://doi.org/10.1515/itit-2014-1061

Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413–423. http://doi.org/10.1007/s11192-016-1910-9

Haustein, Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2013). Coverage and adoption of altmetrics sources in the bibliometric community. In *Proceedings of the 14th International Society of Scientometrics and Informatics Conference*, 15-19 July 2013, Vienna, Austria.

Haustein, S., Bowman, T. D., & Costas, R. (2015). Communities of attention around journal papers: who is tweeting about scientific publications? In *Social Media & Society 2015 International Conference* (pp. 4–5), Toronto, Canada.

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting Altmetrics: Viewing Acts on Social Media through the Lens of Citation and Social Theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication* (pp. 372–406). Berlin, Boston: De Gruyter. http://doi.org/10.1515/9783110308464-022

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, *67*(1), 232–238. http://doi.org/10.1002/asi.23456

Haustein, S., & Larivière, V. (2014). Mendeley as a source of global readership by students and postdocs? Evaluating article usage by academic status. In *In 35th International Association of Technological University Libraries (IATUL), 2-5 June 2014, Aalto University, Helsinki, Finland. Helsinki* (pp. 1–10). Helsinki, Finland.

Haustein, S., & Larivière, V. (2015). The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects. In *Incentives and Performance* (pp. 121–139). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-09785-5_8

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, *65*(4), 656–669. http://doi.org/10.1002/asi.23101

Holmberg, K. (2017). *Different types of Altmetrics. Policy Support Facility (PSF), Mutual Learning Exercise (MLE) on Open Science: Altmetrics and Rewards, European Commission.*

Jamali, H. R., Nicholas, D., & Herman, E. (2016). Scholarly reputation in the digital age and the role of emerging platforms and mechanisms. *Research Evaluation*, *25*(1), 37–49. http://doi.org/10.1093/reseval/rvv032

Jamali, H. R., Nicholas, D., Watkinson, A., Herman, E., Tenopir, C., Levine, K., … Nichols, F. (2014). How scholars implement trust in their reading, citing and publishing activities: Geographical differences. *Library and Information Science Research*, *36*(3–4), 192–202. http://doi.org/10.1016/j.lisr.2014.08.002

Jobmann, A., Hoffmann, C. P., Künne, S., Peters, I., Schmitz, J., & Wollnik-korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie – Praxis Und Forschung*, *3*(1), 1–19.

Ke, Q., Ahn, Y., & Sugimoto, C. R. (2016). A Systematic Identification and Analysis of Scientists on Twitter. *PLOS ONE*, *12*(4), 1–28. http://doi.org/10.1371/journal.pone.0175368

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. http://doi.org/10.1002/asi.23369

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media reserch methods* (pp. 13–26). Sage Publications Inc.

Meijer, I. (2012). Societal returns of scientific research How can we measure it ? Societal returns of scientific research. *"New Perspectives on Enduring Research Questions in University-Society Interaction?", INTERACT-UNI EU-SPRI Conference*, (i), 1–13.

Moed, H. F. (2015). Altmetrics as traces of the computerization of the research process, *8*(93), 4–17. Retrieved from http://arxiv.org/abs/1510.05131

Moed, H. F. (2017). *Applied evaluative informetrics*. Springer.

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. http://doi.org/10.1002/asi.23286

National Information Standards Organization. (2016). *Outputs of the NISO Alternative Assessment Metrics Project*. Retrieved from: http://www.niso.org/apps/group_public/download.php/17091/NISO RP-25-2016 Outputs of the NISO Alternative Assessment Project.pdf

Nicholas, D., Jamali, H. R., Watkinson, A., Herman, E., Tenopir, C., Volentine, R., … Levine, K. (2015). Do Younger Researchers Assess Trustworthiness Differently when Deciding what to Read and Cite and where to Publish? *International Journal of Knowledge Content Development {&} Technology*, *5*(2), 45–63. http://doi.org/10.5865/IJKCT.2015.5.2.045

Nielsen, M. (2011). Definitions of Open Science? Open Knowledge. Retrieved April 12, 2018, from http://openscience.org/an-informal-definition-of-openscience/

Orduña-Malea, E., Martín-Martín, A., & López-Cózar, E. D. (2016). ResearchGate como fuente de evaluación científica: desvelando sus aplicaciones bibliométricas. *El Profesional de La Información (EPI)*, *25*(2), 303–310.

Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159. http://doi.org/10.1038/493159a

Priem, J. (2010). I like the term #articlelevelmetrics, but it fails to imply *diversity* of measures. Lately, I'm liking #altmetrics. Retrieved from: https://twitter.com/jasonpriem/status/25844968813

Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multidimensional indicators of performance* (pp. 263–287). Cambridge, MA: MIT Press.

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

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved from http://arxiv.org/abs/1203.4745

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). altmetrics : a manifesto. Retrieved from http://altmetrics.org/manifesto/

Puschmann, C., & Mahrt, M. (2012). Scholarly blogging: A new form of publishing or science journalism 2.0? In M. Tokar, S. Beurskens, M. Keuneke, I. Mahrt, C. Peters, C. Puschmann, & T. van Treeck (Eds.), *Science and the Internet* (pp. 171–182). Düsseldorf: Düsseldorf University Press.

Ràfols, I., van Leeuwen, T. N., & Robinson-García, N. (2017). How to make altmetrics useful in societal impact assessments: shifting from citation to interaction approaches. Retrieved from:
http://blogs.lse.ac.uk/impactofsocialsciences/2017/03/23/how-to-make-altmetrics-useful-in-societal-impact-assessments-shifting-from-citation-to-interaction-approaches/

Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, *12*(8), e0183551. http://doi.org/10.1371/journal.pone.0183551

Rousseau, R., & Ye, F. (2013). A multi-metric approach for research evaluations. *Chinese Science Bulletin*, *58*(26), 3288–3290.

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195.
http://doi.org/10.1087/20110306

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research Blogs and the Discussion of Scholarly Information. *PLoS ONE*, *7*(5), e35869. http://doi.org/10.1371/journal.pone.0035869

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, *65*(5), 1018–1027. http://doi.org/10.1002/asi.23037

Shuai, X., Pepe, A., & Bollen, J. (2012). How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *PLoS ONE*, *7*(11), e47523. http://doi.org/10.1371/journal.pone.0047523

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. http://doi.org/10.1002/asi.23833

Taraborelli, D. (2008). Soft peer review: Social software and distributed scientific evaluation. In *Proceedings of the 2008 International Conference on the Design of Cooperative Systems,* (pp. 99–110).

Taylor, M. (2013). The Challenges of Measuring Social Impact Using Altmetrics - Research Trends. *Research Trends*, (33), 1–6.
Retrieved from http://www.researchtrends.com/issue-33-june-2013/the-challenges-of-measuring-social-impact-using-altmetrics/

Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, *113*(3), 1721–1731. http://doi.org/10.1007/s11192-017-2557-x

Thelwall, M. (2018). Does Microsoft Academic find early citations? *Scientometrics*, *114*(1), 325–334. http://doi.org/10.1007/s11192-017-2558-9

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, *8*(5), e64841. http://doi.org/10.1371/journal.pone.0064841

Thelwall, M., & Kousha, K. (2014). Academia.edu: Social network or Academic Network? *Journal of the Association for Information Science and Technology*, *65*(4), 721–731. http://doi.org/10.1002/asi.23038

Thelwall, M., & Kousha, K. (2015a). ResearchGate: Disseminating, communicating, and measuring Scholarship? *Journal of the Association for Information Science and Technology*, *66*(5), 876–889. http://doi.org/10.1002/asi.23236

Thelwall, M., & Kousha, K. (2015b). Web indicators for research evaluation. Part 2: Social media metrics. *El Profesional de La Información*, *24*(5), 607. http://doi.org/10.3145/epi.2015.sep.09

Thelwall, M., & Kousha, K. (2017). ResearchGate versus Google Scholar: Which finds more early citations? *Scientometrics*. http://doi.org/10.1007/s11192-017-2400-4

Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics*, *12*(1), 237–248.
http://doi.org/10.1016/j.joi.2018.01.008

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting Links to Academic Articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics,* 1-8.

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? In *In Proceedings of the 2015 International Society for Scientometrics and Informetrics (pp. 95–100). Istanbul, Turkey.*

Tsou, A., Bowman, T. D., Sugimoto, T., Larivière, V., & Sugimoto, C. R. (2016). Self-presentation in scholarly profiles: Characteristics of images and perceptions of professionalism and attractiveness on academic social networking site…. *First Monday*, *21*(4), 1–30.

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: exploring and combining bibliometrics in search of scientific excellence. *Scientometrics*, *57*(2), 257–280. http://doi.org/10.1023/A:1024141819302

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, *512*(7513), 126–129. http://doi.org/10.1038/512126a

Waltman, L., & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology*, *65*(3), 433–445. http://doi.org/10.1002/asi

Wang, X., Fang, Z., Li, Q., & Guo, X. (2016). The Poor Altmetric Performance of Publications Authored by Researchers in Mainland China. *Frontiers in Research Metrics and Analytics*, *1*(October), 1–8. http://doi.org/10.3389/frma.2016.00008

Wass, J. (2017). Crossref Event Data: A new, open, underlying data set for altmetrics. In *The 2017 Altmetrics Workshop*: t*he Dependencies of Altmetrics*, 26 September 2017, Toronto, Canada.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., … Hill, J. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. http://doi.org/10.13140/RG.2.1.4929.1363

Wilsdon, J. R., & Al., E. (2017). *Next-generation metrics : Responsible metrics and evaluation for open*. Brussels. Retrieved from http://doi.org/10.2777/337729

Wouters, P. (2014). A key challenge: the evaluation gap. *The Citation Culture*, 2012–2014. Retrieved from https://citationculture.wordpress.com/2014/08/28/a-key-challenge-the-evaluation-gap/

Wouters, P., & Costas, R. (2012). Users , narcissism and control – tracking the impact of scholarly publications in the 21 st century. Utrecht: SURFfoundation. Retrieved from http://www.surf.nl/nl/publicaties/Documents/Users narcissism and control.pdf

Wouters, P., Thelwall, M., Kousha, K., Waltman, L., De Rijcke, S., Rushforth, A., & Franssen, T. (2015). The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management). *Hefce*, 188. http://doi.org/10.13140/RG.2.1.5066.3520

Wouters, P., Zahedi, Z., & Costas, R. (2018). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of Quantitative Science and Technology Research*. Springer.

Zahedi, Z., & Costas, R. (2017). How visible are the research of different countries on WoS and twitter? An analysis of global vs. local reach of WoS publications on Twitter. In *Proceedings of the 16th of International Conference on Scientometrics and Informetrics,* Wuhan, China.

Zahedi, Z., Costas, R., & Wouters, P. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? . In *PLoS ALM Workshop, 7-9 October, 2013, San Francisco, US.* Retrieved from https://openaccess.leidenuniv.nl/handle/1887/23579

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers ? Study of 1000 PLOS ONE publications using the PLOS ALM , Mendeley and Altmetric . com APIs. In *altmetrics14: expanding impacts and metrics An ACM Web Science Conference 2014 Altmetrics Workshop, 23-26 June, Indiana University, Indiana, USA.* (pp. 5–6). Indiana: Figshare. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1041821

CHAPTER 2

How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications[1]

[1] This chapter is based on:
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, 101(2), 1491–1513. http://doi.org/ 10.1007/s11192-014-1264-0

Abstract

In this chapter an analysis of the presence and possibilities of altmetrics for bibliometric and performance analysis is carried out. Using the web based tool Impact Story, we collected metrics for 20,000 random publications from the Web of Science. We studied both the presence and distribution of altmetrics in the set of publications, across fields, document types and over publication years, as well as the extent to which altmetrics correlate with citation indicators. The main result of the study is that the altmetrics source that provides the most metrics is Mendeley, with metrics on readerships for 62.6% of all the publications studied, other sources only provide marginal information. In terms of relation with citations, a moderate spearman correlation (r=0.49) has been found between Mendeley readership counts and citation indicators. Other possibilities and limitations of these indicators are discussed and future research lines are outlined.

Keywords
Altmetrics, Impact Story, Citation indicators, Research evaluation

2.1. Introduction

Citation based metrics and peer review have a long tradition and are widely applied in research evaluation. Citation analysis is a popular and useful measurement approach in the context of science policy and research management. Citations are usually considered as a proxy for 'scientific impact' (Moed, 2005). However, citations are not free of limitations (Mac Roberts & Mac Robert, 1989, Nicolaisen, 2007), they only measure a limited aspect of quality (i.e., the impact on others' scientific publication) (Martin & Irvin, 1983; Bornmann & Leydesdorff, 2013), their actual meaning has been broadly debated (Wouters, 1999) and they also pose technical and conceptual limitations (Seglen, 1997; Bordons, Fernandez, & Gomez 2002). On the other hand, peer review or peer assessment is also an important instrument and is often regarded as gold standard in assessing the quality of research (Thelwall, 2004; Moed, 2005; Butler & Macalister, 2011;Taylor, 2011; Hicks & Melkers, 2012), but it has its own limitations and biases as well (Moed, 2007; Benos et al., 2007). Moreover, both citations and peer review are considered mostly as partial indicators of "scientific impact" (Martin & Irvin, 1983) and also no single metric can sufficiently reveal the full impact of research (Bollen et al., 2009). Given these limitations, the combination of peer review with "multi-metric approach" is proposed as necessary for research evaluation (Rousseau & Ye, 2013) in the line of the "informed peer review" idea suggested by Nederhof & van Raan (1987).

However, the shortcomings of these more traditional approaches in assessing research have led to the suggestion of new metrics that could inform "new, broader and faster" measures of impact aimed at complementing traditional citation metrics (Priem, Piwowar & Hemminger, 2012). This proposal of using and applying so-called 'alternative indicators' in assessing scientific impact has entered the scientific debate, and these new metrics are expected not only to overcome some of the limitations of the previous approaches but also to provide new insights in research evaluation (Priem & Hemminger, 2010; Galligan & Dyas-Correia, 2013; Bornmann, 2013).

These alternative metrics refer to more "unconventional" measures for evaluation of research (Torres-Salinas, Cabezas-Clavijo & Jimenez-Contreras, 2013), including metrics such as usage data analysis (download and view counts) (Blecic, 1999; Duy & Vaughan, 2006; Rowlands & Nicholas, 2007; Bollen, Van de Sompel, & Rodriguez, 2008; Shuai, Pepe & Bollen, 2012); web citation and link analyses (Smith, 1999; Thelwall, 2001; Vaughan & Shaw, 2003; Thelwall, 2008; Thelwall, 2012) or social web analysis (Haustein, 2010). The importance of the web as a rich source for measuring impact of scientific publications and its potentials to cover the inadequacies of current metrics in research evaluation have been also acknowledged in these previous studies. For instance, the scholarly evidence of use of publications found on web are seen as complimentary to citation metrics, also as predictors of later citations (Brody, Harnad & Carr, 2006) and being of relevance for fields with less citations (Armbruster, 2007). In this sense, the more traditional metrics based on citations, although widely used and applied in research evaluation, are unable to measure the online impact of scientific literature (for example via Facebook, Twitter, reference managers, blogs or wikis) and also lack the ability of measuring the impact of scholarly outputs other than journal articles or conference proceedings, ignoring other outputs such as datasets, software, slides, blog posts, etc. Thus, researchers who publish online and in formats different than journal articles do not really benefit from citation based data metrics.

The rise of these new metrics has been framed with the proposition of the so-called "altmetrics" or social media metrics introduced in 2010 by Priem and colleagues (Priem et al., 2010) as an alternative way of measuring broader research impacts in social web via different tools (Priem, Piwowar & Hemminger, 2012; Priem, et al. 2012). More specifically, altmetrics covers mentions of scientific outputs in social media, news media and reference management tools. This development of the concept of altmetrics has been accompanied by a growth in the diversity of tools that aim to track 'real-time'[1] impact of scientific outputs by exploring the shares, likes, comments, reviews, discussions, bookmarks, saves, tweets and mentions of scientific publications and sources in social media (Wouters & Costas, 2012). Among these tools we find F1000 (http://f1000.com), PloS Article-Level-Metrics (ALM) (http://article-level-metrics.plos.org/), Altmetric.com (www.altmetric.com), Plum Analytics (www.plumanalytics.com/), Impact Story[1] (www.impactstory.org/), CiteULike (www.citeulike.org/), and Mendeley (www.mendeley.com/). These web based tools capture and track a wide range of researcher's outputs by aggregating altmetrics data across a wide variety of sources. In the next section, we summarize the previous studies on altmetrics that have made use of these tools.

2.2. Background

The study of altmetrics is in its early stage but some work has already been done. The features of altmetrics tools in general (Zhang, 2012) and their validation as a sources of impact assessment has been investigated in some studies. For example, Li & Thelwall & Giustini (2012) studied the strengths, weaknesses and usefulness of two reference management tools for research evaluation. Their findings showed that compared to CiteULike, Mendeley seems to be more promising for future research evaluation. Wouters & Costas (2012) compared features of 16 web based tools and investigated their potentials for impact measurement for real research evaluation purposes. They concluded that although these new tools are promising for research assessment, due to their current limitations and restrictions, they seem to be more useful for self-analysis than for systematic impact measurement at different levels of aggregation.

Shuai, Pepe & Bollen (2012) examined the reactions of scholars to the newly submitted preprints in *arXiv.org*, showing that social media may be an important factor in determining the scientific impact of an article. The analysis of social reference management tools compared to citations has been broadly studied in the field, particularly the comparison of citations and readership counts in Mendeley, in most of the cases showing a moderate and significant correlation between the two metrics (Henning, 2010; Priem, Piwowar, &

[1] Previously known as Total Impact, we use IS in this study to refer to Impact Story. For a review of tools for tracking scientific impact see Wouters & Costas (2012).

Hemminger, 2012; Li & Thelwall & Giustini, 2012; Bar-Ilan, 2012; Zahedi, Costas & Wouters, 2013; Schlögl, et al. 2013; Thelwall et al.,2013; Haustein et al., 2013). Also weak correlations between users' tags and bookmarks (as indicators) of journal usage, perception and citations observed for physical journals (Haustein & Siebenlist, 2011) have been reported. For the case of F1000, it has been found that both Mendeley user counts and F1000 article factors (FFas) in Genomics and Genetics papers correlate with citations and they are associated with Journal Impact Factors (Li & Thelwall, 2012).

Some other studies have focused on whether altmetrics can be used as predictor of citations. For example, in the case of F1000, it has been found that recommendations have a relatively lower predictive power in indicating high citedness as compared to journal citation scores (Waltman & Costas, 2013). It has been also suggested that at the paper level, tweets can predict highly cited papers within the first 3 days of publication (Eysenbach, 2011) although these results have been criticized by Davis (2012) and more research should delve into this point. Moreover, most of the articles that received blog citations close to their publication time are more highly cited than articles without such blog citations (Shema, Bar-Ilan & Thelwall, 2013).

Previous studies mentioned above used altmetrics as a new data source and investigated the association between altmetrics and citation impact. Most of these studies were based on journals such as *Nature* & *Science (*Li, Thelwall, & Giustini, 2012); *JASIST (*Bar-Ilan, 2012), Information System Journal (Schlögl, et, al., 2013); articles published by bibliometrics and scientometrics community (Bar-Ilan et al., 2012; Haustein et al., 2013), *PLoS* and other medical and biomedical journals in PubMed (Priem, Piwowar & Hemminger, 2012; Thelwall et al., 2013; Haustein et al., 2013).

However, to the best of our knowledge, little has been done to date to investigate the presence of altmetrics across various scientific fields and also for relatively ample periods of time. This study is thus one of the first in analyzing a relatively large sample of publications belonging to different fields, document types and publication years. This paper builds upon Wouters & Costas (2012) and Zahedi, Costas & Wouters (2013).

Our main objective in this paper is to present an exploratory analysis of altmetrics data retrieved through Impact Story focusing on the relationship of altmetrics with citations across publications from different fields of science, social sciences and humanities. For this, we examine the extent to which papers have altmetrics obtained through different data sources retrieved via Impact Story and the relationships between altmetrics and citations for these papers. In exploring these issues, we pursue the two following research questions:

1) What is the presence and distributions of Impact Story altmetrics across document types, subject fields and publication years for the studied sample?

2) Is there any relationship between Impact Story-retrieved altmetrics and citation indicators for the studied sample? In other words, to what extent do the Impact Story altmetrics correlate with citation indicators?

2.3. Research methodology

In this study, we have focused on Impact Story (IS). Although still at an early stage ('beta version'), IS is currently one of the most popular web based tools with some potentials for research assessment purposes (Wouters & Costas, 2012). IS aggregates "impact data from many sources and displays it in a single report making it quick and easy to view the impact of a wide range of research output" (http://impactstory.org/faq). It takes as input different types of publication identifiers (e.g., DOIs, URLs PubMed ids, etc.). These are run through different external services to collect the metrics associated with a given 'artifact' (e.g., a publication). A final web based report is created by IS which shows the impact of the 'artifacts' according to a variety of metrics such as the number of readers, bookmarks, tweets, mentions, shares, views, downloads, blog posts and citations in Mendeley, CiteULike, Twitter, Wikipedia, Figshare, Dryad, Scienceseeker, PubMed and Scopus[2].

For this study, we collected a random sample of 20,000 publications with DOIs (published between 2005 and 2011) from all the disciplines covered by the Web of Science (WoS). Publications were randomly collected by using the "*NEW ID ()*" SQL command (Forta 2008, p. 193).The altmetrics data collection was performed during the last week of April 2013. The altmetrics data were gathered automatically via the Impact Story REST API[3], then the responses provided on search requests using DOI's were downloaded. Using this API we could download the altmetric data faster (one request per 18 seconds) compared to the manual data collection we did for the previous study[4]. The files were downloaded per API search request separately in Java Script Object Notations (JSON) format on the basis of individual DOI's and parsed by using the additional JAVA library from within the SAS software[5]. Finally, the data was transformed into a Comma Separated value (CSV) format and matched back with the CWTS in-house version of the Web of Science on the DOIs to be able to add other bibliometric data to them. The final list of publications resulted in 19,772 DOIs (out of 20,000) after matching[6]. Based on this table, we studied the distribution of altmetrics across subject

[2] For a full list see http://impactstory.org/faq

[3] A REpresentational State Transfer (REST)(ful) API (Application Programming Interface) used to make a request using GET (DOIs) and collect the required response from impact Story.

[4] In the previous study, the data collection was performed manually directly through the web interface of IS. Manually, IS allowed collecting altmetrics for 100 DOIs per search and maximum 2000 DOIs search per day in order to avoid swamping the limits of its API, for details see Zahedi, Costas & Wouters (2013).

[5] The additional functionality from the "proc groovy" which is a java development environment added to SAS (Statistical analysis Systems) environment for parsing and reading the JSON format and returning the data as an object.

[6] From IS one DOI was missing. We also found that 301 DOIs were wrong in WoS (including extra characters that made them unmatchable, therefore excluded from the analysis). Also 61 original DOIs from WOS pointed to 134 different WOS publications (i.e., being duplicated DOIs). This means that 74 publications were duplicates. Given the fact that there was no systematic way to determine which one was the correct one (i.e., the one that actually received the altmetrics), we included all of them in the analysis with the same altmetrics score resulted in: 20000-1-301+74=19772 final publications. All in all, this process showed that only 1.8% of the initial DOIs randomly selected had some problems, thus indicating that a DOI is a convenient publication identifier although not free of limitations (i.e., errors in DOI data entry, technical errors when resolving DOIs via API and also the existence of multiple publication identifiers in the data sources, resulted in some errors in the full collection of altmetrics for these publications).

fields, document types and publication years. Citation indicators were calculated and the final files were imported in IBM SPSS Statistics 21 for further statistical analysis. In order to test the validity of our sample set we compared the distribution of publications across major fields of science in our sample with that of the whole Web of Science database (Figure 1) in the same period and only those publications with a DOI. As it can be seen, the distribution of publications of our sample basically resembles the distribution of publications in the whole WOS database, so we can consider that our sample is representative of the multidisciplinarity of the database.

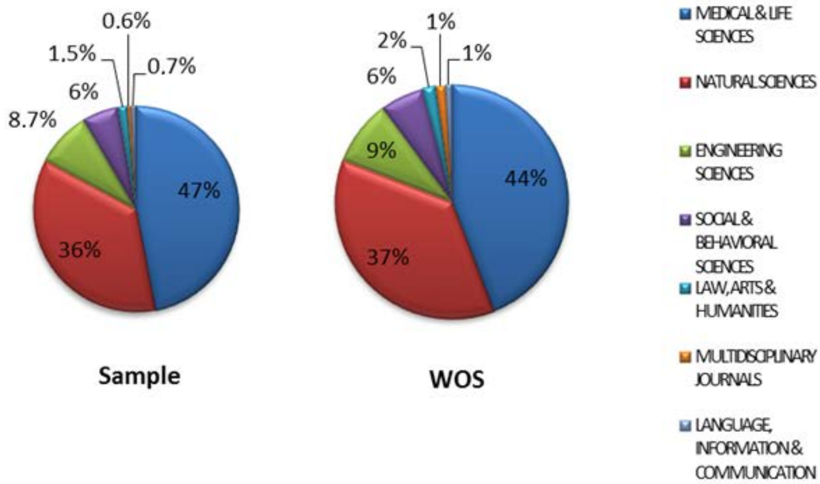


Figure 1. Distribution of publications by major fields of science: sample vs. whole database.

2.4. Results and main findings

In the first place, we present the result of our exploratory analysis of the presence of IS altmetrics over the 19,772 WOS publications published between 2005-2011. Then, we examine the extent to which papers are represented in the data sources both in general and also across document types, subject fields and publication years. Finally, the relationships (correlation) between IS altmetrics and citations for these papers are compared.

2.4.1.Presence of IS altmetrics by data sources

In our sample, the presence of IS altmetrics across publications is different from each data source. Out of 19,722 publications, 12,380 (62.6%) papers have at least one reader[7] in Mendeley, 324 (1.6%) papers have at least one tweet in Twitter, 289 (1.4%) papers have at least one mention in Wikipedia, 72 (0.3%) papers have at least one bookmark in Delicious and 7413 (37.4%) papers have at least one citation in PubMed. Only 1 paper in the sample has metrics from PLoS ALM[8]. Based on this preliminary test, we decided to exclude some of the metrics from our study: PlosAlm indicators due to their low frequency as they are only available for the PLoS journals thus their presence in our sample is negligible and PubMed-based citations because they are limited only to the Health Sciences and they refer to

[7] It means that publications without any metrics were left out of the analysis.
[8] This was the only PLOS paper captured by our sample.

citations, which we will calculate directly based on the Web of Science. We also decided to sum the metrics coming from Twitter ("Topsy tweets" and "Topsy influential tweets") given their relatively low frequency. As a result, in the current study, the data from Mendeley, Wikipedia, Twitter and Delicious were analyzed. Table 1 shows the number and percentages of papers with and without IS altmetrics sorted by % of papers with metrics (excluding the PLOS ALM and PubMed metrics). Based on Table 1, our main finding is that, for this sample, the major source for altmetrics is Mendeley, with metrics on readerships for 62.6% of all the publications studied. But for other data sources (Twitter, Wikipedia and Delicious), the presence of metrics across publications is very low, with more than 98% of the papers without metrics. Thus, it is clear that their potential use for the assessment of the impact of scientific publications is still rather limited, particularly when considering a multi-year and multidisciplinary dataset as the one here studied.

Table 1. Presence of IS altmetrics from data sources.

Data Source	*papers with metrics*	*%*	*papers without metrics*	*%*
Mendeley	12380	**62.6**	7392	37.3
Twitter	324	1.6	19448	98.3
Wikipedia	289	1.4	19483	98.6
Delicious	72	.3	19700	99.7

2.4.2. Presence of IS altmetrics across document types

Regarding document type, out of 19772 publications, there are 16740 (84.7%) articles, 944 (4.7%) review papers, 487 (2.4%) letters and 1601(8%) non-citable[9] items in the sample. Table 2 indicates the coverage of the sampled publications with document types across each data sources. According to Table 2, 81.1% (766) of the review papers, 66.3% (11094) of articles, 25.1% of letters and 24.9% (398) of non-citable in the sample have been saved (read) in the Mendeley. In Twitter, 3.4% (32) of the review papers, 1.9% (30) of non-citable items, 1.5% (255) of articles and 1.4% (7) of letters have tweets. In the case of Wikipedia, 4.6% (43) of the review papers, 1.4% (230) of articles and less than 1% of other document types (letters and non-citable) are mentioned at least once in Wikipedia. Therefore, Mendeley has the highest coverage of all data sources in this sample, (81.1% of the review papers and 66.3% of articles in the sample are covered by Mendeley).

Table 2. Coverage of publications with different document types by different data sources.

Doc. Type	**pub**	%	**Mendeley**	%	**Twitter**	%	**Wikipedia**	%	**Delicious**	%
article	16740	84.7	11094	66.3	255	1.5	230	1.4	56	0.3
review	944	4.7	766	81.1	32	3.4	43	4.6	7	0.7
letter	487	2.4	122	25.1	7	1.4	4	0.8	3	0.6
Non citable	1601	8.0	398	24.9	30	1.9	12	0.7	6	0.4
Total	**19772**	100	**12380**	**62.6**	**324**	**1.6**	**289**	**1.4**	**72**	**0.3**

[9] Non-citable document type corresponds to all WoS document types other than article, letter and review (e.g., book reviews, editorial materials, etc.).

We also studied the total numbers of Mendeley readers, tweets, mentions and bookmarks for each document types covered in the sample (i.e., not only the number of publications with metrics, but the frequency of these metrics). Table 3 shows the result of the total sum and the average number of altmetrics scores per document types provided by the different data sources. Based on both table 3 and figure 2, in general, articles have the highest values of numbers of readers, tweets and bookmarks (more than 77.5% of all altmetrics scores are to articles), followed[10] by review papers, non-citables and letters (less than 18% of the altmetrics scores are to the other types) in all data sources. But considering the average metrics per publications[11], it can be seen that, Mendeley accumulate the most metrics per all document types than all other data sources. Also, in Mendeley, review papers have attracted the most readers per publications (on average there are ~14 readers per review paper) than all other data sources.

Table 3. Distribution of IS altmetrics per document types in different data sources.

Doc Type	pub	Mendeley Readers	%	Avg	Tweets	%	Avg	Wikipedia Mentions	%	Avg	Delicious Bookmarks	%	Avg
Article	16740	82553	83.3	4.9	3020	94.5	0.18	292	77.5	0.02	213	87.3	0.01
Review	944	12730	12.9	13.4	78	2.4	0.08	68	18.0	0.07	7	2.9	0.01
Letter	487	466	0.5	0.9	21	0.7	0.04	4	1.1	0.01	10	4.1	0.02
Non-citable	1601	3301	3.3	2.0	76	2.4	0.05	13	3.4	0.01	14	5.7	0.01
Total	**19772**	99050	100	5.0	3195	100	0.16	377	100	0.02	244	100	0.01

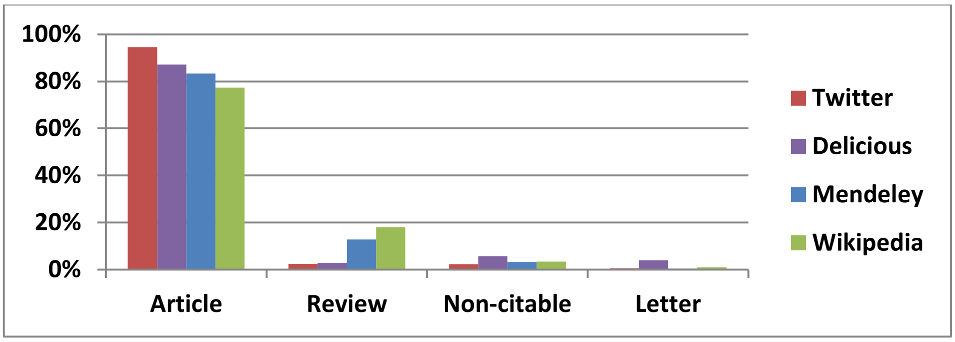


Figure 2. Distribution of IS altmetrics across document types

2.4.3. Presence of IS altmetrics across NOWT Subject fields

For this analysis, we used the NOWT (High) classification which has 7 major disciplines developed by CWTS[12]. Table 4 shows the percentage of publications having at least one metrics (i.e., papers with at least one reader in Mendeley, once bookmarked in Delicious, once tweeted, or once mentioned in Wikipedia) across those major disciplines[13]. According to the

[10] in Delicious, articles, non-citables, letters and review papers have the highest number of metrics orderly.

[11] Average metrics per publications calculated by dividing the total numbers of metrics from each data source by total number of publications in the sample. For example, in Mendeley, average number of readers per publication equals to 99050/19772=~5

[12] In the previous study, we used the NOWT (Medium) with 14 subject fileds. For more details see: http://nowt.merit.unu.edu/docs/NOWT-WTI_2010.pdf

[13] Here publications can belong to multiple subject categories.

results, Multidisciplinary publications ranked the highest in all data sources. The major source for altmetrics data in our sample is Mendeley with the highest proportion for Multidisciplinary fields, which include journals such as Nature, Science or PNAS. 80% of the publications in this field, 73% of the publications from Medical & Life Sciences[14] and 68% of the publications from Social & Behavioural Sciences have at least one Mendeley reader. Among the other data sources, Multidisciplinary publications ranked the highest as well but with lower presence of publications with metrics. Regarding the top three fields with the highest percentage of altmetrics, Wikipedia has similar pattern as Mendeley: 7% of the publications from Multidisciplinary field, 2% of the publications from Medical & Life Sciences and 2% of the publications from Social & Behavioural Sciences have at least one mention in Wikipedia. In Twitter, 7% of the publications from Multidisciplinary field, 3% of the publications from Social & Behavioural Sciences and 2% of publications from Medical & Life Sciences are the top three fields that have at least one tweet. In Delicious, only 1% of the publications from Multidisciplinary field, Language, Information & Communication and Social & Behavioural Sciences have at least one bookmark while other fields have less than 1% altmetrics.

Table 4. Coverage of publications with different NOWT subject fields by different data sources.

NOWT High Subject Categories	Total number of publications	%	Mendeley	%	Wikipedia	%	Twitter	%	Delicious	%
MULTIDISCIPLINARY JOURNALS	216	47	172	80	15	7	16	7	3	1
MEDICAL & LIFE SCIENCES	15637	36	11353	73	284	2	301	2	67	0.4
SOCIAL & BEHAVIORAL SCIENCES	1878	6	1268	68	32	2	58	3	11	1
NATURAL SCIENCES	11935	8.7	6554	55	103	1	123	1	34	0.3
ENGINEERING SCIENCES	2885	0.6	1558	54	7	0.2	9	0.3	2	0.1
LANGUAGE. INFORMATION & COMMUNICATION	241	0.7	123	51	2	1	1	0.4	3	1
LAW ARTS HUMANITIES	488	1.5	190	39	8	2	7	1	0	0
		100								

Again, the total scores of Mendeley readers, tweets, mentions and bookmarks for each discipline in the sample have been calculated. Figure 3 shows that the distributions of IS altmetrics across different subject fields is uneven. Both Medical & Life and Natural Sciences received the highest proportion of altmetrics in all data sources. In general in all data sources, more than 30% of altmetrics accumulated by publications from Medical & Life Sciences and

[14] According to the Global Research Report by Mendeley (http://www.mendeley.com/global-research-report/), coverage of Mendeley in different subjects are as follows: the highest coverage are by publications from Biological Science & Medicine (31%), followed by Physical Sciences and Maths (16%), Engineering & Materials Science (13%), Computer & Information Science (10%), Psychology, Linguistics & Education(10%), Business Administration, Economics & Operation Research (8%), Law & Other Social Sciences (7%) and Philosophy, Arts & Literature & other Humanities (5%)

more than 23% of altmetrics are to publications from the fields of Natural Sciences. Other fields, each received less than 10% of total altmetrics. Comparing the different data sources in terms of the proportion of altmetrics across fields, different patterns arise: Medical & Life Sciences fields proportionally attracted the most attention in Wikipedia, followed by Mendeley, Twitter and Delicious while in case of Natural Sciences, Delicious, Twitter, Mendeley and Wikipedia, proportionally got the most attention orderly; moreover, for Mendeley, both Social & Behavioural and Engineering Sciences, proportionally, received the highest attention than all other fields.

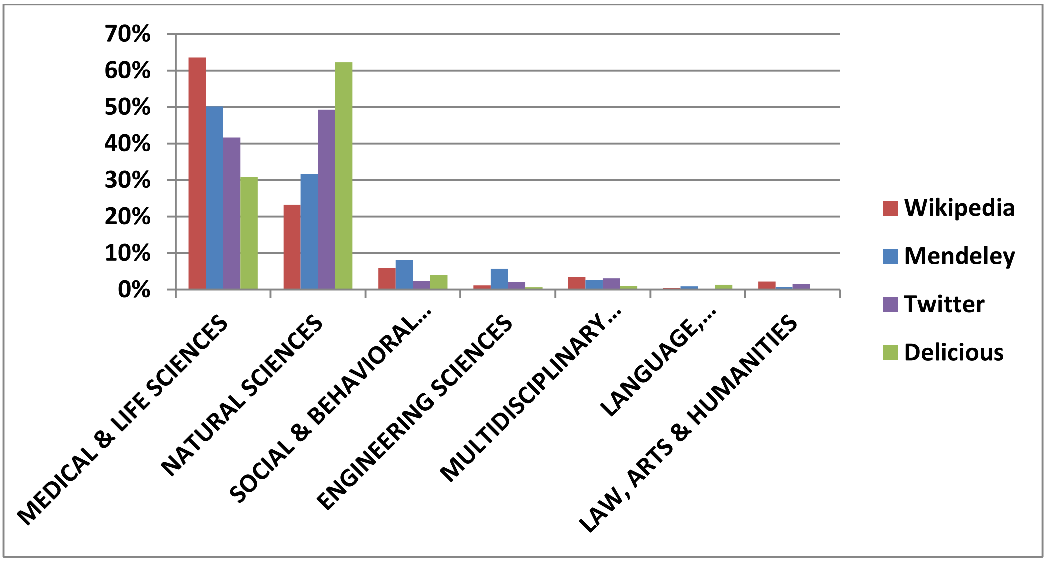


Figure 3. Distribution of IS altmetrics across NOWT subject fields.

Table 5. Distribution of IS altmetrics per NOWT subject fields in different data sources.

NOWT Subject category	Mendeley Readers		Wikipedia Mentions		Delicious Bookmarks		Tweets	
MEDICAL & LIFE SCIENCES	86347	50%	371	64%	92	31%	1958	42 %
NATURAL SCIENCES	54481	32%	136	23%	186	62%	2317	49 %
SOCIAL & BEHAVIORAL SCIENCES	14102	8%	35	6%	12	4%	112	2%
ENGINEERING SCIENCES	9800	6%	7	1%	2	1%	100	2%
MULTIDISCIPLINARY JOURNALS	4521	3%	20	3%	3	1%	144	3%
LANGUAGE, INFORMATION & COMMUNICATION	1492	0.9 %	2	0.3 %	4	1%	1	0%
LAW, ARTS & HUMANITIES	1297	0.8 %	13	2%	0	0%	72	2%
		100		100		100		100

2.4.4. Comparison of Citations per Papers (CPP) and Readerships per Papers (RPP) across fields

Although measuring the impact of scholarly publications in social media is very important, it is not yet clear for what purposes scholarly publications are mentioned in social media and reference management tools such as Mendeley, in social bookmark manager such as Delicious, in Wikipedia and Twitter by different users/scholars, and particularly it is not clear if these mentions can be considered as measures of any type of "impact" of the publications. In case of Mendeley, it is assumed that publications are saved in users' libraries for immediate or later reading and possibly also future citation.

In any case, it is important to know how many altmetrics vs. citations each publication received and what are the different pattern across different subject fields. Due to the fact that not all of scholarly publications are covered equally by citation databases and also the existence of disciplinary differences in terms of citations, which vary a lot between fields, it is interesting to study both the proportion of altmetrics vs citations per publications to see which fields can benefit from having more density of altmetrics scores (i.e., altmetrics scores per paper) than citation density. Since Twitter, Wikipedia and Delicious showed an overall very low presence per paper, we focus here only on Mendeley. Both the average number of Mendeley readerships per papers (RPP) and WoS citations per papers (CPP) across different NOWT subject fields were calculated and analyzed (Figure 4). For calculating the citations (excluding self-citations), we used a variable citation window from the year of publication to 2012. Also a variable "readership window" was considered for Mendeley, counting readerships from the publication year of the paper until the last week of April 2013. In this analysis we have also included publications without any metrics (citations or Mendeley readers). The result (Figure 4 sorted by RPP) shows that in general, Multidisciplinary journals have the highest values of both RPP and CPP; and Law, Arts & Humanities have the lowest values. For fields such as Multidisciplinary journals, Medical & Life Sciences, Natural and Engineering Sciences, the value of CPP is higher than RPP, while for fields such as Social & Behavioural Sciences, Language, Information & Communication and Law, Arts & Humanities, RPP outperforms CPP. The latter is an interesting result that might suggest the relevance of Mendeley for the study of Social Sciences and Humanities publications, which are often not very well represented by citations (Nederhof 2006). In order to further test the differences between RPP and CPP, we extended the same type of analysis for all 248 WOS individual subject categories, resulting that 167 out of 248 WOS subject categories have higher CPP values than RPP values. Most of the fields with higher values of CPP vs. RPP are from the Sciences (145), 18 from the Social Sciences and 4 from the Art and Humanities. On the other hand, 72 fields presented higher RPP than CPP scores (among them 31 are from Social Sciences, 27 from Science and 13 from Art and Humanities)[15]. Therefore, we can conclude that citations are more dominant than readerships particularly in the fields of the Sciences (which are also the fields with the highest coverage in citation databases); while on the other hand, many sub-fields from the Social Sciences and Art

[15] For 9 fields (8 fields from Art and Humanities and 1 field from Science) CPP and RPP scores were exactly the same.

and humanities received proportionally more readerships per paper than citations per paper. This could be seen as a possibility for these fields with lower coverage in citation databases (such as WoS) to benefit from Mendeley in terms of having more readership impact than citation impact, although this needs further explorations.

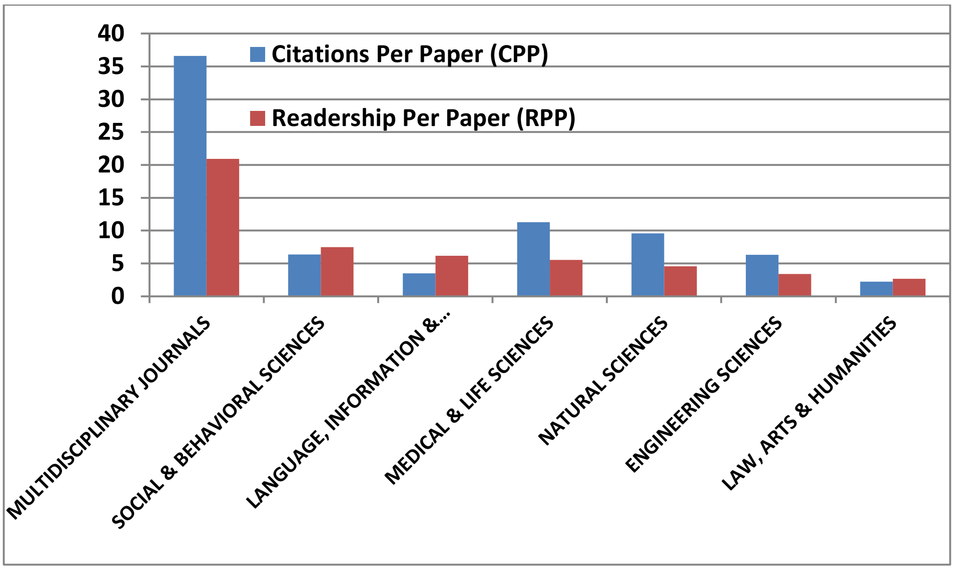


Figure 4. Comparing CPP and RPP in Mendeley across Subject Fields.

2.4.5. Trend analysis of IS altmetrics across publication years

Table 6 shows the trend analysis of number and share of publications in the sample by altmetrics sources. Regarding the publication years, the share of publications ranges from 10% in the year 2005 to 18% in the year 2011. The coverage of different sources is also shown in the table. In our sample, Mendeley has its peak in its proportion of publications with some readers in 2009 (66%) and the lowest point in 2011 (57%), although the total number of publications with some Mendeley readerships has increased during the whole period, with the exception of 2011 when there is a small drop compared to 2010. Twitter has its highest peak in 2011 (4%) and its lowest values in the early years (around 1% between 2005-2009). Wikipedia mentions are for 2% of all the publications published between 2005 to 2008 and 1% of all the publications published between 2010 and 2011. For Delicious, the highest peak is for the years 2007 and 2011 and the lowest one for the year 2005, also publications from 2008, 2009 and 2010 have the same presence in Delicious. All in all, it seems that Twitter and Delicious tend to cover the more recent publications better than the older ones although the values are in general very low.

Table 6. Coverage of publications with different publication years by different data sources.

Pub year	p		Mendeley		Wikipedia		Delicious		Twitter	
2005	2006	10%	1263	63%	39	2%	3	0.1%	17	1%
2006	2405	12%	1491	62%	58	2%	4	0.2%	6	0.2%
2007	2682	14%	1702	63%	41	2%	13	0.5%	16	1%
2008	2858	14%	1799	63%	46	2%	11	0.4%	34	1%
2009	3039	15%	2001	66%	43	1%	12	0.4%	31	1%
2010	3228	16%	2099	65%	37	1%	13	0.4%	62	2%
2011	3548	18%	2020	57%	25	1%	16	0.5%	158	4%

The presence of overall altmetrics scores (i.e., not only publications with altmetrics, but their total counting) has been also calculated in order to know its trend over time. According to Table 7, this is quite different across different data sources. For example, for Wikipedia and Mendeley, publications from the years 2006 and 2009, accumulated most of the mentions (20%) and readerships (17%) respectively. In the case of Mendeley and Wikipedia we noticed a decrease in the amount of altmetrics in the last two years.

Both in Delicious and in Twitter, publications from the year 2008 received the highest proportion of altmetrics. In case of Delicious, 50% of bookmarks and in case of Twitter, 34% of tweets are to publications published in 2008. Comparing the amount of altmetrics in each year across different data sources shows that in this sample, both the oldest and the most recent publications in Twitter have the most altmetrics (tweets) (26% of tweets are to publications from the year 2005[16] and 2011 respectively) and also the recent publications (2009-2010) have the most altmetrics (readerships) in Mendeley (figure 5).

Table 7. Distribution of IS Altmetrics across publication year.

Pub year	p		Mendeley		Wikipedia		Delicious		Twitter	
2005	2006	10%	10814	11%	48	13%	51	21%	835	26%
2006	2405	12%	12658	13%	77	20%	4	2%	20	1%
2007	2682	14%	13739	14%	58	15%	14	6%	102	3%
2008	2858	14%	14299	14%	67	18%	122	50%	1072	34%
2009	3039	15%	16922	17%	50	13%	21	9%	145	5%
2010	3228	16%	16305	16%	43	11%	14	6%	198	6%
2011	3548	18%	14239	14%	34	9%	18	7%	823	26%
		100%		100%		100%		100%		100%

[16] In 2005, the two most tweeted papers are from the field of Physics, they received more than half of the total tweets in this year (472 tweets), thus showing a strong skewed distribution.

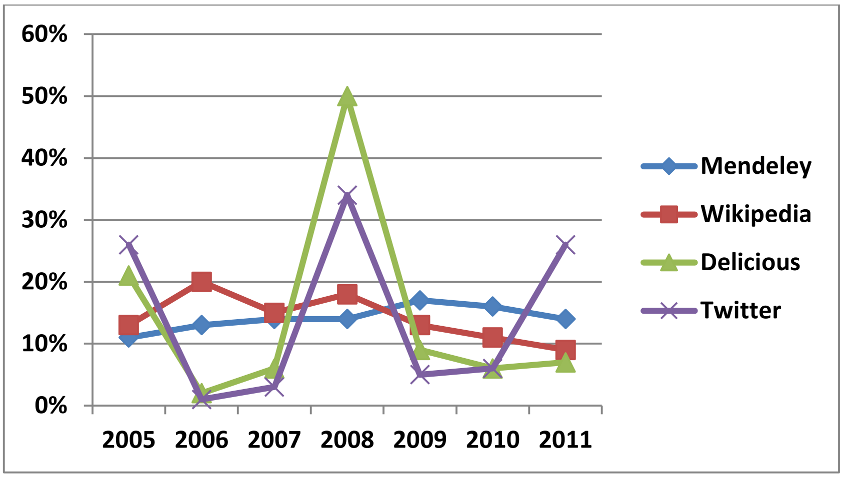


Figure 5. Distribution of IS altmetrics across publication years.

2.4.6. Relationships between IS altmetrics and citation indicators

In this section we study more thoroughly the relationship between the IS altmetrics and citation indicators. Following the CWTS standard calculation of indicators (cf. Waltman et al., 2011), we calculated for all the publications the following citation indicators: *Citation Score* (CS), that is, number of citations per publications; *Normalized Citation Score* (NCS), that is, number of citations per publications, with a normalization for fields differences and publication year; *Journal Citation Score* (JCS), that is the average number of citations received by all publications in that journal of a publication; and *Normalized Journal Score* (NJS), that is, the average number of citations received by all publications in that journal normalized by fields differences and publication year. For the calculation of the impact indicators, as explained before, we used a variable citation window (i.e., citations up to 2012) excluding self-citations. The result of the factor analysis, the correlation analysis and impact of publications with and without altmetrics will be presented in the next sections.

2.4.7. Factor analysis of IS altmetrics and bibliometrics indicators

An exploratory factor analysis has been performed using SPSS version 21 in order to know more about the underlying structure, relationship among the variables and the dimension of variables (Table 8). Principal Component Analysis (PCA) revealed the presence of 2 main components or dimensions with eigenvalues exceeding 1, explaining 58% of the total variance. The first dimension is dominated by bibliometric indicators. Mendeley readerships and Wikipedia mentions are also included in this dimension; although Mendeley readership counts has the highest loadings in this dimension of the two indicators. The second dimension is more related to social media metrics, showing that Twitter and Delicious are strongly correlated. These results suggest that the variables in each group may represent similar concepts.

Table 8. Factor analysis of the variables.

Rotated Component Matrix a		
	Component	
	1	2
CS	**.837**	*.005*
NCS	**.752**	*.009*
JS	**.745**	*-.011*
NJS	**.720**	*-.015*
Mendeley	**.680**	*.008*
Wikipedia	**.297**	*.009*
Delicious	*.003*	**.954**
Twitter	*.004*	**.954**
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Loadings higher than .1 are shown. 58% of total variance explained.		
a. Rotation converged in 4 iterations.		

2.4.8. Correlations between IS altmetrics and bibliometrics indicators

In order to overcome the technical limitation of SPSS for calculating Spearman correlation for large datasets[17], first, rankings of variables computed using Data>rank cases and then Pearson correlation performed on the ranked variables; this method provides the spearman correlation of the original variables. Table 9 shows the result of the correlation analysis among the different altmetrics data source and citation and journal citation scores and their 95% confidence intervals (calculated using the Bootstrapping technique implemented in SPSS). According to this table, citation indicators are more correlated between them than with altmetrics. In general, direct citations indicators (i.e., CS and NCS) correlate better among them than with indicators of journal impact (JS and NJS), although the correlations between the two groups are fairly high. Mendeley is correlated with Wikipedia (r=.08) and Twitter is correlated with Delicious (r=.12), this is in line with the result of the factor analysis but the correlation values are very low. Compared to citation indicators, Mendeley has the highest correlation score with citations (moderate correlation of r=0.49) among all the altmetrics sources. The other altmetric sources show very weak or negligible correlation with citation indicators.

[17] Calculating Spearman correlation analysis in SPSS for large datasets gives this error: "Too many cases for the available storage", for overcoming this limitation, we followed the process we mentioned in the text. For more details see: http://www-01.ibm.com/support/docview.wss?uid=swg21476714

Table 9. Correlation analysis of the rank values of variables.

	NCS	JS	NJS	Mendeley	Wikipedia	Delicious	Twitter
CS	.886 (.882-.89)	.762 (.756-.769)	.557 (.547-.567)	.497 (.485-.508)	.094 (.08-.108)	.011 (-.005-.027)	.025 (.01-.039)
NCS		.528 (.516-.538)	.6 (.59-.609)	.467 (.455-.478)	.074 (.059-.087)	.019 (.002-.035)	.054 (.037-.068)
JS			.711 (.702-.718)	.44 (.428-.452)	.09 (.075-.105)	-.003 (-.018-.012)	-.003 (-.018-..011)
NJS				.427 (.415-.439)	.058 (.044-.072)	.012 (-.005-.028)	.039 (.023-.053)
Mendeley					.083 (.067-.099)	.031 (.015-.047)	.07 (.055-.084)
Wikipedia						.021 (-.001-.049)	.056 (.025-.087)
Delicious							.125 (.073-.185)

2.4.9. Impact of publications with/without altmetrics

In this section, we study the differences in impact between publications with and without altmetrics. The main idea is to see whether publications with altmetrics tend to have more citation impact than those without altmetrics. Table 10 presents the bibliometric indicators and their 95% confidence intervals (calculated using the Bootstrapping technique implemented in SPSS). For instance, according to the median values it can be observed that publications with metrics have in general higher citation scores compared to those without metrics in all data sources (although, in some cases, the confidence intervals show some overlapping, thus the claim of the higher impact for these cases is less strong and probably more influenced by outliers).

Table 10. Comparison of NCS and NJS of the publications with and without altmetrics.

		With Metrics				Without Metrics			
		CS	**JS**	**NCS**	**NJS**	**CS**	**JS**	**NCS**	**NJS**
Mendeley	**N**	12380	12380	12380	12380	7392	7392	7392	7392
	Median	5	6.53	0.72	1.02	1	1.76	0.10	0.53
Confidence Interval	**Lower**	4	6.4	0.69	1.01	0.5	1.67	0.08	0.51
	Upper	5	6.69	0.74	1.04	1	1.89	0.12	0.55
Wikipedia	**N**	289	289	289	289	19483	19483	19483	19483
	Median	12	13.87	1.18	1.18	2	4.43	0.47	0.86
Confidence Interval	**Lower**	9	11.91	0.97	1.07	2	4.32	0.46	0.85
	Upper	14	15.2	1.35	1.31	3	4.57	0.49	0.87
Twitter	**N**	324	324	324	324	19448	19448	19448	19448
	Median	4	3.6	1	1.1	3	4.53	0.47	0.86
Confidence Interval	**Lower**	3	3.1	0.86	0.97	2	4.39	0.46	0.85
	Upper	5	4.74	1.29	1.27	3	4.62	0.49	0.87

Delicious	**N**	72	72	72	72	19700	19700	19700	19700
	Median	3	3.99	0.89	1.07	3	4.52	0.48	0.86
Confidence Interval	**Lower**	2	2.34	0.52	0.76	2	4.38	0.46	0.85
	Upper	6	5.55	1.57	1.33	3	4.62	0.49	0.87

Focusing on the number of Mendeley readers per publication and considering their impact as measured by the NCS and NJS, we can see how publications tend to increase in citation impact as the number of readerships increases (Figure 6). The effect is quite strong, especially for the average number of citations per publication but this is less prominent for the NJS indicator. The same result found by Waltman & Costas (2013) for relationship between recommendations from F1000, citations and journal impact. In their study, they found that on average, publications with more recommendations also have higher citation and journal impact.

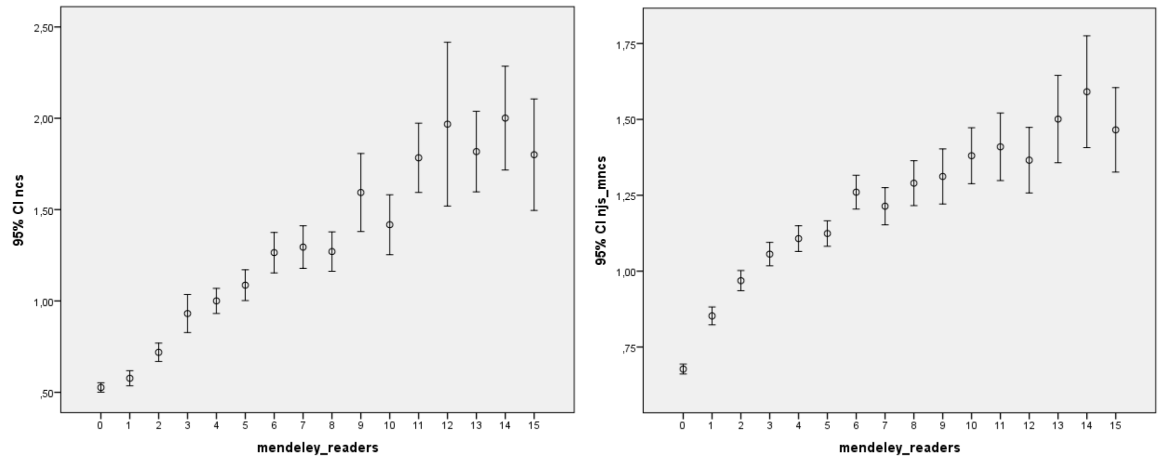


Figure 6. Relation between number of Mendeley readerships and citation and journal impact.

2.5. Discussion and conclusions

In this paper we have used Impact Story[18] for gathering altmetrics for a set of randomly sampled publications. IS is an interesting open source for collecting altmetrics, however, we also see some important limitations[19] particularly regarding the speed and capacity of data collection and formatting of the data. We detect different results comparing our current results with those presented in our previous study (Zahedi, Costas & Wouters, 2013) mostly due to the different methodology of data collection (manually vs. automatically) and collecting the data at different points in time as it happened between our two studies, where in the first one, Mendeley was only presented in around 37% of the publications[20] and now in more than

[18] Impact Story, was in an initial stage of development (i.e., in a 'Beta' version) at the moment of development of this study.
[19] For current limitations of IS see: http://impactstory.org/faq#toc_3_11
[20] The time interval between the first and the second data collection was 6 months and data collection done manually versus the second one which done automatically using RESTAPI calls.

60% [21]. This situation also points to the need for the tools to be transparent in how their data are collected and their limitations. This means that an important natural future step will be the proper assessment of the validity of the data retrieved via different altmetrics data sources (as it has been done for example for Google Scholar – cf. Delgado López-Cózar et al., 2012). This validation of the quality, reliability and robustness of the altmetrics tools is essential in order to be able to apply altmetrics for serious research assessment purposes. For these tools to be fully incorporated in regular research assessment processes, they need to meet the necessary requirements for data quality, transparency and indicator reliability and validity as emphasized by Wouters & Costas (2012) in their study of altmetric tools. Moreover, the results of this study are based on the WOS covered publications; hence, it is important to keep in mind the restrictions of this database with regards to its coverage of some fields, language and publication formats (Moed, 2009; Van Raan, Van Leeuwen & Visser, 2011; Archambault & Larivière, 2006; Torres-Salinas, Cabezas-Clavijo & Jimenez-Contreras, 2013).

All in all, given the exploratory nature and the fact that basically the same results have been found with the two data collections, we can assume that our results are robust and valid for our purposes. In general, our study shows that Mendeley is the major and more useful source for altmetrics data. Mendeley has the highest coverage and proportion of altmetrics compared to Twitter, Wikipedia and Delicious for the studied publications. Out of 19,772 publications a total 12380 cases (62.6%) had at least one reader in Mendeley. Previous studies also showed that Mendeley is the most exhaustive altmetrics data source (Bar-Ilan et al., 2012, Priem et al., 2012) mostly for the publications from Library and Information Science field: 97.2% coverage for JASIST articles published between 2001 and 2011 (Bar-Ilan, 2012); 82% coverage for articles published by researchers in Scientometrics (Bar-Ilan et al., 2012); and 82% of bibliometrics literature (Haustein et al., 2013), for Multidisciplinary journals such as Nature and Science (94% and 93% of articles published these journals in 2007) (Li, Thelwall, & Giustini, 2012); and more than 80% of PLoS ONE publications (Priem, et al. , 2012) covered by Mendeley. In terms of document type, review papers and articles were proportionally the most read, shared, liked or bookmarked format compared to non-citable items and letters across all data sources. Multidisciplinary fields (i.e., the field where journals such as *Nature*, *Science* or the *PNAS* are included) are the most present in all altmetrics data sources but concerning the distribution of altmetrics across different fields, more than 30% of altmetrics accumulated by publications from Medical & Life Sciences and more than 23% of altmetrics are to publications from the fields of Natural Sciences. Comparing both proportion and distribution of IS altmetrics across different fields among different data sources shows different patterns, particularly in Mendeley, both Social & Behavioural and Engineering Sciences, have proportionally received the highest attention compared to all other fields**.** Considering citations and readerships per publication, Multidisciplinary journals have the highest and Law, Arts & Humanities have the lowest density of both citations and readerships

[21] Reasons for these differences can be the changes/improvements in the Identification of publications by Mendeley (e.g., by merging version of the same paper, Identifying more DOIs, increment in the number of users in Mendeley, etc.

per publications. However, according to our observation, there is a higher density of readerships per paper than citations per papers in several fields of the Social Sciences and Humanities. This finding suggests that Mendeley readership counts could have some added value in supporting the evaluation and analysis of these fields, which have been traditionally worse represented by citation indicators (cf. Nederhof, 2006). Another explanation for those fields with lower proportion of readers than citations could be the fact that Mendeley is relatively new and not yet widely used and adopted among all scholars from all the disciplines. Besides, differences in citation and readership behaviors and practices among fields could also explain these differences. In any case, this is an aspect that needs further analysis.

Our trend analysis shows that particularly publications with Mendeley readerships have increased over time, although there is a slight decrease in the number of readerships and proportion of publications with Mendeley readers for the last two years. The most plausible explanation for this is that the accumulation of readers takes some time. To the best of our knowledge there is no information on the ‘readership history’ of publications (besides the fact that readerships could conceptually decrease as the users delete or change their libraries) and so far we don’t have results on the readerships pace. This means that we don’t know when a paper in a given year has obtained its peak in readerships. It is highly likely, that although faster than citations, the accumulation of readerships for publications also takes some time, and this is the reason why for the most recent publications, the number of readers is slower as compared to those older publications that have had more time to accumulate readerships. Future research should also focus on disentangling this aspect.

The Spearman correlation of Mendeley readerships with citation impact indicators showed moderate correlations (r=.49) between the two variables which is also found in other previous studies (Bar-Ilan, 2012; Priem et al., 2012). This indicates that reading and citing are related activities, although still different activities that would be worthwhile to explore. According to the result of comparing the impact of publications with and without altmetrics with their citation scores, it can be also concluded that in general, publications with more altmetrics also tend to have both higher direct citations and are published in journals of higher impact. The issue about the potential predictability of citations through altmetric scores will be explored in follow-up research.

Finally, although citations and altmetrics (particularly Mendeley readerships) exhibit a moderate positive relationship, it is not yet clear what the quality of the altmetrics data is and neither what kind of dimension of impact they could represent. Since altmetrics is still in its infancy, at the moment, we don’t yet have a clear definition of the possible meanings of altmetric scores. In other words, the key question of what altmetrics mean is still unanswered. From this perspective, it is also necessary to know the motivations behind using these data sources, for example in case of Mendeley: what does it reflect when an item is saved/added by several users to their libraries? Also, what does it mean that an item is mentioned in Wikipedia, CiteULike, Twitter and any other social media platform? Does it refer to the same or different dimension compared to citation? In the same line, besides studying to what extent different publications are presented in Mendeley and other social media tools and their relations with citation impact, we need to study for what purposes and why these platforms are exactly used by different scholars. Moreover, research about the quality and reliability of the altmetric data retrieved by the different altmetrics providers is still necessary before any interpretation and potential real uses for these data and indicators are developed. This information in combination with the assessment of the validity and reliability of altmetrics data and tools will shed more light on the meanings of altmetrics and can help to unravel the hidden dimensions of altmetrics in future studies.

2.6. Acknowledgement

This study is the extended version of our research in progress paper presented at the 14th International Society of Scientometrics & Informetrics Conference, 15-19 July, 2013, Vienna, Austria. We thank the Impact Story team for their support in working with the Impact Story API. This work is partially supported by the EU FP7 ACUMEN project (Grant agreement: 266632). The authors would like to thank Erik Van Wijk from CWTS for his great help in managing altmetrics data. The authors also acknowledge the useful suggestions of Ludo Waltman from CWTS and the fruitful comments of the anonymous referees of the journal.

2.7. References

Archambault, É., & Larivière, V. (2006). The limits of bibliometrics for the analysis of the social sciences and humanities literature. International Social Science Council: World social sciences report 2010: Knowledge divides. Paris: UNESCO, 251–254.

Armbruster, C. (2007). Access, usage and citation metrics: what function for digital libraries and repositories in research evaluation? Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1088453

Bar-Ilan, J. 2012. JASIST@ Mendeley. Presented at ACM Web Science Conference Workshop on Altmetrics, Evanston, IL, 21 June 2012. Retrieved from: http://altmetrics.org/altmetrics12/bar-ilan/

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

Benos, D. J., Bashari, E., Chaves, J. M., et al. (2007). The ups and downs of peer review. Advances in Physiology Education, 31:145-152.

Blecic, D. (1999). Measurements of Journal Use: An Analysis of the Correlations Between Three Methods. Bulletin of the Medical Library Association, 87: 20–25

Bollen, J., Van de Sompel, H., & Rodriguez, M. A. (2008). Towards usage-based impact metrics. In Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL), New York, USA.

Bollen J, Van de Sompel H, Hagberg A, Chute R (2009). A Principal Component Analysis of 39 Scientific Impact Measures. PLoS ONE, 4(6): e6022. https://doi.org/10.1371/journal.pone.0006022

Bordons, M, Fernandez, MT & Gomez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. Scientometrics, 53: 195–206.

Bornmann, L. (2013). Is there currently a scientific revolution in Scientometrics? Journal of the American Society for Information Science & Technology, 65(3). https://doi.org/10.1002/asi.23073

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: a comparative study using data from InCites and F1000. Journal of Informetrics, 7(2), 286-291.

Brody, T., Harnad, S. and Carr, L. (2006), Earlier Web usage statistics as predictors of later citation impact. Journal of the American Society for Information Science, 57: 1060–1072.

Butler, L., & McAllister, I. (2011). Evaluating university research performance using metrics. European Political Science, 10(1), 44-58.

Davis, P. M. (2012). Tweets, and Our Obsession with Alt Metrics. The Scholarly Kitchen. Retrieved fromhttp://scholarlykitchen.sspnet.org/2012/01/04/tweets-and-our-obsession-with-alt-metrics/

Delgado Lopez-Cozar, Emilio, Robinson-Garcia, Nicolas & Torres Salinas, Daniel (2012). Manipulating Google Scholar Citations and Google Scholar Metrics: simple, easy and tempting. Retrieved from: http://arxiv.org/abs/1212.0638

Duy, J. & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. The Journal of Academic Librarianship, 32(5): 512-517.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. Journal of Medical Internet Research, 13(4), e123.

Forta, B. (2008). Sams Teach Yourself SQL in 10 Minutes. USA: Sams Publishing.

Galligan, F., & Dyas-Correia, S. (2013). Altmetrics: Rethinking the Way We Measure. Serials Review, 39(1), 56–61.

Haustein, S. (2010). Multidimensional journal evaluation. In Proceedings of the 11th International Conference on Science and Technology Indicators (pp. 120–122), Leiden, the Netherlands.

Haustein S, Siebenlist T (2011) Applying social bookmarking data to evaluate journal usage. Journal of Informetrics, 5: 446–457.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2013). Coverage and adoption of altmetrics sources in the bibliometric community. In *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference*, 13-19 July 2013, Vienna, Austria.

Henning, V. (2010). The top 10 journal articles published in 2009 by readership on Mendeley. Retrieved from: http://www.mendeley.com/blog/academic-features/the-top-10-journal-articles-published-in-2009-by-readership-on-mendeley/.

Hicks, D. & Melkers, J. (2012). "Bibliometrics as a Tool for Research Evaluation" Handbook on the Theory and Practice of Program Evaluation. Ed. Al Link & Nick Vornatas. Edward Elgar. Retrieved from: http://works.bepress.com/diana_hicks/31

Li, X., & Thelwall, M. (2012). F1000, Mendeley and Traditional Bibliometric Indicators. In: Proceedings of the 17th International Conference on Science and Technology Indicators, Montréal, Canada. pp. 451–551.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. Scientometrics, 91(2), 461–471.

MacRoberts, M. H. & MacRoberts, B. R. (1989), Problems of citation analysis: A critical review. Journal of the American Society for Information Science, 40: 342–349.

Martin B R, Irvine J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. Research Policy, 12:61–90

Moed, H.F. (2005). Citation analysis in research evaluation. Berlin/ Heidelberg/New York: Springer.

Moed, H.F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. Science and Public Policy, 34(8), 575–583.

Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. Archivum immunologiae et therapiae experimentalis, 57(1), 13–8.

Nederhof, A.J., & Van Raan, A. F. J. (1987). Peer review and bibliometric indicators of scientific performance: a comparison of cum laude doctorates with ordinary doctorates in physics. Scientometrics, 11 (5-6): 333-350

Nederhof, A.J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. Scientometrics, 66 (1), 81–100.

Nicolaisen, J. (2007). Citation Analysis. Annual Review of Information Science and Technology, 41, 609–641.

Priem, J., Hemminger, B.H., (2010) Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday 15. Retrieved from: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874/2570

Priem, J., Taraborelli, D., Groth, P., and Neylon, C. (2010). Altmetrics: a manifesto. Retrieved from: http://altmetrics.org/manifesto/

Priem, J., Piwowar, H., & Hemminger, B. H. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. ArXiv: 1203.4745v1

Rousseau, R., & Ye, F. (2013). A multi-metric approach for research evaluation. Chinese Science Bulletin, 10–12.

Rowlands, I & Nicholas, D. (2007). The missing link: journal usage metrics, Aslib Proceedings, 59(3), 222 – 228.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs. citation vs. readership data: the case of an information systems journals. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (eds.), Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, Austria (pp. 626-634). Wien: Facultas Verlags und Buchhandels AG.

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

Shema, H., Bar-Ilan, J., & Thelwall, M. (2013). Do blog citations correlate with a higher number of future citations ? Research blogs as a potential source for alternative metrics. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (eds.), Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, Austria (pp. 604-611). Wien: Facultas Verlags und Buchhandels AG.

Shuai, X, Pepe, A, Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: article downloads Twitter mentions, and citations. Retrieved from: http://arxiv.org/abs/1202.2461v1

Smith, A. G. (1999). A tale of two web spaces; comparing sites using web impact factors. Journal of Documentation, 55(5), 577–592.

Taylor, J. (2011). The assessment of research quality in UK. Universities: peer review or metrics? British Journal of Management, 22 (2), 202–217.

Thelwall, M. (2001). Extracting macroscopic information from web links. Journal of American Society for Information Science and Technology, 52(13), 1157–1168.

Thelwall, M. (2004). Weak benchmarking indicators for formative and semi-evaluative assessment of research. Research Evaluation, 13 (1), 63–68.
Thelwall, M. (2008). Bibliometrics to Webometrics. Journal of Information Science, 34(4), 605-621.
Thelwall, M. (2012). A history of webometrics. Bulletin of the American Society for Information Science and Technology, 38(6), 18–23.
Thelwall, M. (2012). Journal impact evaluation: A webometric perspective, Scientometrics, 92(2), 429-441.
Thelwall M, Haustein S, Larivière V, Sugimoto CR (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLoS ONE, 8(5), e64841.
Torres-Salinas, D., Cabezas-Clavijo, A., Jimenez-Contreras, E. (2013). Altmetrics: New indicators for scientific communication in web 2.0. Comunicar. Retrieved from: http://arxiv.org/ftp/arxiv/papers/1306/1306.6595.pdf
Torres-Salinas, D., Robinson-Garcia, N., Campanario, J. M., & López-Cózar, E. D. (2013). Coverage, field specialisation and the impact of scientific publishers indexed in the Book Citation Index. Online Information Review, 38(1), 24–42.
Van Raan, A. F. J., Van Leeuwen, T. N., & Visser, M. S. (2011). Severe language effect in university rankings: particularly Germany and France are wronged in citation-based rankings. Scientometrics, 88(2), 495–498.
Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations: what is the difference? Journal of the American Society for Information Science and Technology, 54(14), 1313–1322.
Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S., & Van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics, 5(1), 37–47.
Waltman, L., Costas, R. (2013). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.23040
Wouters, P., Costas, R. (2012). Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation. Retrieved from: http://www.surffoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf
Wouters, P. (1999), The Citation Culture, PhD thesis, University of Amsterdam.
Zahedi, Z., Costas, R. & Wouters, P. (2013). How well developed are Altmetrics? Cross disciplinary analysis of the presence of 'alternative metrics' in scientific publications (RIP). In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (eds.), Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, Austria (pp. 876-884). Wien: Facultas Verlags und Buchhandels AG.
Zhang, Y. (2012). Comparison of select reference management tools. Medical reference services quarterly, 31(1), 45–60.

CHAPTER 3

General discussion of data quality challenges in social media metrics: extensive comparison of four major altmetric data aggregators[1]

[1] This chapter is based on:
Zahedi, Z., and Costas, R. (2018). General discussion of data quality challenges in social media metrics: extensive comparison of five major altmetric data aggregators. PloS one, 13(5), e0197326. http://doi.org/10.1371/journal.pone.0197326

Abstract

The data collection and reporting approaches of four major altmetric data aggregators are studied. The main aim of this chapter is to understand how differences in social media tracking and data collection methodologies can have effects on the analytical use of altmetric data. For this purpose, discrepancies in the metrics across aggregators have been studied in order to understand how the methodological choices adopted by these aggregators can explain the discrepancies found. Our results show that different forms of accessing the data from diverse social media platforms, together with different approaches of collecting, processing, summarizing, and updating social media metrics cause substantial differences in the data and metrics offered by these aggregators. These results highlight the importance that methodological choices in the tracking, collecting, and reporting of altmetric data can have in the analytical value of the data. Some recommendations for altmetric users and data aggregators are proposed and discussed.

Keywords

Altmetric data; altmetric data aggregators; data quality; social media metrics; social media data collection; recommendations

3.1. Introduction

Altmetrics offer the possibility of studying new forms of interactions between social media users, scholarly objects, and different academic actors. As such, altmetrics or rather social media metrics (Wouters, Zahedi, and Costas, 2018) have paved the way towards the study of the relationships and interactions between social media and scholarly entities in what can be seen as the *social media studies of science* (Costas, 2017). However, for the proper development of this new genre of studies it is critical to understand all possible data quality challenges in the capture of social media events around scholarly objects (Haustein, 2016). Therefore, questions such as *from where*, *when* and *how* social media data has been collected and processed become critical in the development of reliable and replicable social media metrics research. Besides, the existence of different altmetric data aggregators opens the question of how these aggregators are approaching the collection of social media data; and how their different approaches may introduce discrepancies in the results based on their data. The study of social media metrics data quality is a critical and central element in the further development of altmetric research.

3.1.1. Data issues in bibliometrics

The importance of relying on reliable and valid data has always been a bone of contention in bibliometric research. The occasional lack of transparency of some bibliometric databases (e.g., Web of Science, Scopus, Google Scholar), together with the errors and inconsistences found in citations, have been often highlighted in the literature (Halevi, Moed, and Bar-Ilan, 2017; Moed, Bar-Ilan, and Halevi, 2016; Delgado Lopez-Cozar, Robinson-Garcia, and Torres-Salinas, 2013), particularly regarding their potential effect in research evaluation. For instance, errors such as inaccurate cited references and duplicate records in Scopus or Web of Science (Olensky, Schmidt, and van Eck, 2016) have been discussed to have serious consequences in the calculation of citation indicators for journals, individuals, or institutions (Franceschini et al, 2016; Valderrama-Zurián et al., 2015). Given the importance of the data quality of bibliometric data, comparative analyses of bibliometric and citation data sources have proliferated (Jacsó, 2011), often providing recommendations on how to improve the accuracy and quality of bibliometric databases (Halevi, Moed, and Bar-Ilan, 2017).

3.1.2. Data issues in social media metrics data

In the case of social media metrics data sources much less is known about their potential issues regarding their data quality. Different approaches in collecting, processing, reporting, and updating the data have been discussed to largely influence the social media metrics offered by different altmetric aggregators (Zahedi, Fenner, and Costas, 2015; Zahedi, Bowman, and Haustein, 2014). The possibility to track the provenance of the original data is considered an important aspect regarding the verification of the data and metrics provided (Chamberlain, 2013). However, only few studies have systematically compared different altmetric aggregators based on their coverage of publications and calculation of metrics

(Meschede & Siebenlist, 2018; Bar-Ilan and Halevi, 2017; Zahedi, Fenner, and Costas, 2014; Jobmann et al., 2014; Chamberlain, 2013). These previous studies have pointed out that the social media metrics data reported can be influenced by issues related with the different timing of data collection, different sources (of for example blog lists or main stream news) tracked, use of APIs (commercial vs. public), or the choice of publication identifiers (e.g., DOIs, PMIDs) to access and track social media data. Therefore, similar to citation data (Harzing and Alakangas, 2016; Meho and Sugimoto 2009; Bar-Ilan, 2008), it is important to understand how variations in the social media metrics reported by the different aggregators may influence the results obtained.

A more recent study (Ortega, 2017) showed that major altmetric aggregators (Plum Analytics, Altmetric.com, CrossRef Event Data) provide different metrics for the same set of papers. These differences challenge the reliability of social media metrics. Possible solutions could be just to select some specific sources (e.g., those that provide the highest scores or coverage) or even their combination, as it seems to be suggested by Ortega (2017). However, for example the selection of aggregators with higher scores do not necessarily mean better indicators or data. For example, higher scores can be caused by the combination of different recorded actions coming from the same social media source (e.g., by counting under the same indicator Facebook shares, likes and wall posts publications, instead of keeping them as separate metrics) or by aggregating metrics from duplicate records of the same object. Such choices may cause even more unreliable results, since different sources of error could be merged in the same indicator. Hence, from our point of view, it is very important to understand the underlying reasons of the existing differences. Also, it is important to discuss how different methodological and technical choices can influence the metrics provided. From this perspective, we aim at providing a more reasoned discussion of the current challenges of social media metrics data, instead of a mere recollection of who is providing the higher (lower) scores or the description of data issues in altmetric sources.

3.1.3. Altmetric data aggregators

Among the most important altmetric aggregators currently collecting and providing social media metrics we can highlight Altmetric.com, Lagotto, Plum Analytics, and CrossRef Event Data. These altmetric data aggregators offer access to data and metrics related with the online activity and social media interactions between social media users and scholarly objects. We also include here the description of social media metrics (number of readers) obtained from Mendeley.com. Although Mendeley.com is an altmetric data provider and not an altmetric data aggregator (National Information Standards Organization, 2016), it is included in this study in order to compare the results of a direct data collection from Mendeley with that of other aggregators. Thus it is possible to better discuss the potential differences found in Mendeley metrics provided by the different aggregators with a common benchmark (i.e., our own data collection from Mendeley). In order to simplify the terminology throughout the paper we will refer to all of them as aggregators, even if we sometimes refer to Mendeley.com.

Altmetric.com (http://www.altmetric.com/)

Altmetric.com is a Digital Science company founded in 2011 and based in London (United Kingdom). More than 64 million mentions of 9 million research outputs are covered by Altmetric.com database in January 2018[1]. A range of different sources[2] including mentions in policy documents, blogs, mainstream media, online reference managers, and social media tools, etc. are tracked for URLs or scholarly outputs unique identifiers (e.g., PubMed ids, ArXiv ids). Counts for each tracked object (journal articles, datasets, images, reports,) are available via its detail page[3] and the recorded data are available for free for researchers through the Altmetric API with a rate limit.

CrossRef Event Data (www.eventdata.crossref.org/)

Crossref Event Data (CrossRef ED) is a service started in April 2017 and is in its beta version. Event Data[4] collects raw data from a selection of sources[5] such as Wikipedia, Twitter, Reddit, Stack Exchange Network, etc. for CrossRef registered contents. This service connects to some external data sources via its agent for turning the data into 'events' (bookmarks, comments, shares) and provides provenance, context, and links for each event. The resulted events are publically available via an open Event Data API. It is important to highlight that this service doesn't provide metrics but a stream of Events (raw data) that occurred for a given piece of registered content with a DOI (Wass, 2017).

Lagotto open source application (www.lagotto.io/)

Lagotto is an Open Source application started in March 2009 by the Open Access publisher Public Library of Science (PLOS). Lagotto started by providing social media mentions of PloS articles and later also for articles from any other publisher. Lagotto retrieves (version 4.2.1. released on 13 July 2015) data from a wide set of services and sources[6]. The metrics are grouped in different categories of impact (viewed, saved, cited, and recommended) and are available through an open API[7].

Plum Analytics (https://plumanalytics.com/)

Plum Analytics was founded in 2012, acquired by Ebsco in 2014 and by Elsevier in 2017. *Plum Analytics* provides metrics for different research outputs (articles, blog posts, books, source codes, theses/dissertations, videos) via its 'artifact' [object] level page. The metrics are grouped in 5 categories of usage, captures, mentions, social media, and citations. PlumX is a

---

[1] https://www.altmetric.com/about-our-data/how-it-works/
[2] https://www.altmetric.com/about-our-data/our-sources/
[3] https://www.altmetric.com/details/950642
[4] https://www.eventdata.crossref.org/guide/data/about-the-data/#sources-and-agents
[5] https://www.eventdata.crossref.org/guide/data/about-the-data/
[6] http://www.lagotto.io/docs/sources/
[7] http://www.lagotto.io/docs/api/

subscription-based platform and hence no open API is available; however, artifact-level PlumX pages are free and publicly accessible[8].

Mendeley (http://www.mendeley.com/)
Mendeley is a free online reference manager and academic social network founded in 2007 and acquired by Elsevier in 2013. This platform is used by over 6 million users worldwide[9]. It offers 'readership' statistics capturing the number of different Mendeley users that have saved a given publication (together with their academic statuses, countries and disciplines)[10].

3.1.4. Methods of collecting, tracking, and updating social media metrics

Although these aggregators may collect data from similar data sources (e.g., Twitter, Facebook, Wikipedia, Mendeley), it is common that they adopt different methodological approaches when collecting, processing, and reporting the data (Zahedi, Fenner, Costas, 2014). From a conceptual point of view, we argue that there are three main central elements in the systematization of the different methodological approaches adopted by the different altmetric data aggregators:

- *Data collection approaches*. Not all altmetric aggregators track the same document types (books, reviews, articles, datasets, slides), journals, or publishers. They also vary in the social media sources they cover (e.g., Facebook, Twitter, Wikipedia, Mendeley). Aggregators may also use different APIs to access the primary sources (e.g., Altmetric.com, Plum Analytics, and CrossRef ED use GNIP for Twitter data while Lagotto uses Search API).
- *Aggregation and reporting approaches*. Aggregators may differ in recording public vs. private data (e.g., public walls Facebook posts are tracked by Altmetric.com while Lagotto and Plum Analytics also track private posts and shares). Social media metrics may be reported with different degrees of detail, thus moving from mere counts or summaries of events to providing the raw metadata collected (e.g., CrossRef ED provides the raw data collected and no counts, while aggregated data at the output level are displayed in Plum Analytics, Altmetric.com, or Lagotto). Aggregators also differ in the scholarly object identifiers (DOIs, PMIDs, Arxiv IDs) they track. Different data processing approaches may also be used. Some aggregators may choose to aggregate tweets and retweets in one single count (e.g., Lagotto), keep them separate (e.g., Plum Analytics, although in their total count they sum them together), just provide the count of the distinct tweeters

[8] It is not clear to us why Ortega (2017) indicated that "[Plum Analytics] does not permit to retrieve publications searching by DOI". From our experience, through the following URL https://plu.mx/plum/a/?doi= it is possible to explore the metrics provided by PlumX for any given DOI.
[9] https://www.mendeley.com/research-network/community
[10] https://www.mendeley.com/reference-management/stats

around a publication (e.g., Altmetric.com), or the raw metadata of the (re)tweets mentioning the scholarly objects (e.g., CrossRef ED).

- *Updating approaches*. Different criteria to update the social media data (daily, weekly) are also applied by the different altmetric aggregators.

3.1.5. Aim of the study

As explained above, there are indeed similarities and differences on how data aggregators approach the data collection, processing, and update of social media events around different scholarly objects and their identifiers. Given these disparities, there is a critical need of understanding how these differences can cause variations in the nature and characteristics of the social media metrics provided. This understanding is fundamental for the future development of robust and reliable applications, ensuring "transparency", "accuracy", and "replicability" as suggested in the literature (National Information Standards Organization, 2016; Wouters & Costas, 2012). Hence, we aim not just at *identifying the potential discrepancies* but also at *conceptualizing the reasons and implications that these differences may have for further social media metrics research.* The paper is organized as follows. In section 2 the main methodological design of this study is described. Section 3 is structured as follows, a first part (section 3.1) includes a quantitative analytical description of the data discrepancies across aggregators. In a second part (section 3.2), a general discussion of potential reasons for the differences found is presented. Finally, some general conclusions and recommendations for social media metrics researchers and altmetric data aggregators are introduced in section 4.

3.2. Data and Methodology

Publications with a DOI published in PloS ONE[11] (n= 31,437) in 2014 and available in the CWTS in-house version of Web of Science (WoS) database have been considered in this study. The DOIs of these publications were used to collect social media metrics data from the described altmetric aggregators using their APIs or dedicated websites[12]:

- Altmetric.com REST API (http://api.altmetric.com/);
- CrossRef Event Data API (www.eventdata.crossref.org/guide/service/query-api/);
- Lagotto open source application API (www.lagotto.io/docs/api/);
- Plum Analytics (https://plu.mx/plum/a/?doi=[*doi*]);
- Mendeley REST API (http://dev.mendeley.com/).

[11] PloS ONE publications were chosen since they are covered and tracked by all aggregators considered for this study.

[12] Impact Story is not used in this study due to its reliance on Altmetric.com for reporting social media metrics (https://www.altmetric.com/blog/impactstory/). Webometrics Analyst (http://lexiurl.wlv.ac.uk/) is a tool for collecting altmetric and webometric data, but cannot be seen as an altmetric data aggregator in itself, therefore it is not used in this study.

The data collection from all the selected altmetric data aggregators was done in exactly the same date: *2017 June 19th* with the aim of minimizing time effects in the data collection. Altmetric data from *Facebook*, *Twitter*, *Mendeley,* and *Wikipedia* obtained from these aggregators were considered for comparisons. Some descriptive statistics such as the sum and average scores of different metrics, and the coverage of publications (% of publications captured by each altmetric aggregators with at least one metric in each of the tracked sources) have been calculated. The (dis)agreement of the metrics provided among altmetric aggregators have also been studied, and Pearson correlations have been calculated in order to determine the relationship between the metrics provided by the aggregators. Possible reasons for the differences found are summarized and discussed, particularly regarding the further development of research and applications of social media metrics.

3.3. Results

In this section, first, we present the differences across aggregators including discrepancies in the coverage of publications, total counts of all and overlapped publications, and correlation analysis of metrics across aggregators.

3.3.1. Differences across aggregators

3.3.1.1. Coverage of publications

The main results of the coverage of publications with some social media recorded activity are presented in Table 1. Overall, Plum Analytics has the highest coverage (99.9%) of PloS ONE publications, followed by Lagotto (99.8%), Mendeley.com (95.9%), and Altmetric.com (61%). CrossRef ED has the lowest coverage (7.5%) of all PloS ONE publications considered in this study. The coverage of publications per data source is presented in the following columns.

Mendeley coverage
Regarding the coverage of publications with at least one Mendeley reader, Plum Analytics has the highest coverage of Mendeley readerships (96.6%), even higher than Mendeley itself (95.8%), followed by Lagotto (95.8%), and Altmetric.com (60.6%). The substantial lower coverage of Altmetric.com is caused by the data collection policy of this aggregator (see 3.2.1 section). CrossRef ED does not collect Mendeley readership.

Twitter coverage
Altmetric.com exhibits the largest coverage of publications with at least one tweet (57%), followed by Lagotto (31.7%), Plum Analytics (23.9%), and CrossRef ED (1.7%). This lower coverage of tweets by CrossRef ED can be related to the recent start of this service, which is still in its Beta version[13].

[13] The Beta version of CrossRef Event Data released on 6 May 2017 and this means that it is not complete and not all functionality are yet implemented. Based on the information from their website stability of Beta service

Facebook coverage

Regarding publications with some coverage on Facebook, Plum Analytics (16.3%) has the largest coverage followed by Altmetric.com (11.5%), and Lagotto (7.9%) while Crossref ED does not collect any Facebook mentions at this moment.

Wikipedia coverage

Lagotto (5.1%) has the highest share of publications with at least one mention in Wikipedia, followed by Plum Analytics (2.3%), CrossRef ED (2.2%), and Altmetric.com (2%).

Table 1. Coverage (% of DOIs with at least one metric) of PloS ONE DOIs across altmetric aggregators and aggregators and per data sources.

Aggregators	#publications with coverage (% pubs.)	#publications on Mendeley (% pubs. on Mendeley)	#publications on Twitter (% pubs. on Twitter)	#publications on Facebook (%pubs. on Facebook)	#publications on Wikipedia (%pubs on Wikipedia)
Altmetric.com	19,185	19,073	17,926	3,623	639
	(61)	*(60.6)*	*(57.0)*	*(11.5)*	*(2.0)*
CrossRef ED	2364	N/A	555	N/A	716
	(7.5)	N/A	*(1.7)*	N/A	*(2.2)*
Lagotto	31,398	30,117	9,973	2,497	1,615
	(99.8)	*(95.8)*	*(31.7)*	*(7.9)*	*(5.1)*
Mendeley.com	30,154	30,124	N/A	N/A	N/A
	(95.9)	*(95.8)*	N/A	N/A	N/A
Plum Analytics	31,418	30,389	7,526	5,149	747
	(99.9)	*(96.6)*	*(23.9)*	*(16.3)*	*(2.3)*

(N/A: metrics not available in the platform)

3.3.1.2. Total counts

In Table 2 the total sum of counts for all publications in each of the studied indicators is presented. Mendeley readership counts include the sum of all readership counts reported by each of the altmetric aggregators. For Altmetric.com the total number of (re)tweets per publication recorded by this source has been calculated by ourselves. The aggregator does not provide per se the total number of (re)tweets a publication has received but the distinct number of tweeters that have (re)tweeted the publication. For CrossRef ED, the count of (re)tweets, number of distinct (re)tweeters, and the count of Wikipedia mentions are calculated by ourselves based on the raw data provided in their JSON files. Plum Analytics already reports the sum of all (re)tweets, Facebook, and Wikipedia counts. It also provides a breakdown by tweets and retweets. The total counts of (re)tweets provided by Plum Analytics is used in this study.

According to Table 2, Lagotto and Plum Analytics provide the highest counts of Mendeley readership (tMR) outperforming the counts provided by Mendeley itself. In terms of Twitter counts (tTW), Altmetric.com reports the highest counts of tweets while CrossRef ED presents

is not guaranteed because this service is still in development and hence there could be some bugs that should be reported, this is explained in more details here: www.eventdata.crossref.org/guide/index.html the lowest values. Plum Analytics provides the highest value of Facebook counts (tFB) and CrossRef ED provides the highest value of Wikipedia counts (tW).

Table 2. Statistics (sum [t] and mean [m] scores) of altmetric counts across aggregators and per data source.

Aggregators nP=31,437	**tMR**	**mMR**	**tTW**	**mTW**	**tFB**	**mFB**	**tW**	**mW**
Altmetric.com	491,630	15.6	164,919 (143,471)	5.2 (4.5)	22,627	0.7	1,060	0.0
CrossRef ED	N/A	N/A	2,912 (2,359)	0.1 (0.07)	N/A	N/A	11,221	0.4
Lagotto	679,898	21.6	104,840	3.3	67,073	2.1	4,683	0.1
Mendeley.com	653,283	20.8	N/A	N/A	N/A	N/A	N/A	N/A
Plum Analytics	671,834	21.4	76,113	2.4	275,122	8.8	1,135	0.0

nP= number of Publication; t=sum score; m=mean score; MR=Mendeley readership counts, TW=(re)tweets, FB=Facebook counts, W= Wikipedia mentions, N/A=metrics not available in the platform, values in parentheses refer to statistics of distinct tweeters (Twitter users) - only for Altmetric.com and CrossRef ED.

3.3.1.3. Counts of overlapped publications

Tables 1 and 2 show that there are indeed differences in the coverage and counts provided by the aggregators. Thus, it is important to delve into the main possible reasons behind these differences. In order to do so, we first explore the level of (dis)agreement in the values of those publications that are covered by the same pairs of altmetric aggregators (i.e., overlapped publications between aggregators).

- *Mendeley readership counts*

A total of 19,073 publications (60.6% of the total) are covered by both Altmetric.com and Mendeley. Of these, 18,613 publications (97.9%) have exactly the same number of readership counts (Table 3). This suggests a strong agreement between Altmetric.com Mendeley data and our method to extract data from Mendeley (as described in Section 2). A total of 153 publications (0.8%) have higher scores recorded in Altmetric.com, while 249 publications (1.3%) have lower scores in Altmetric.com than in Mendeley.

In the case of Lagotto, a total of 30,117 publications (95.8%) are covered in both Lagotto and Mendeley, of which 14,416 publications (47.9%) have exactly the same readership score as reported by Mendeley. In contrast, 13,974 publications (46.4%) have higher scores in Lagotto and 1,727 publications (5.7%) have lower scores in Lagotto than in Mendeley. This suggests a relatively weaker agreement between our method to query Mendeley and that of Lagotto.

For Plum Analytics, although it exhibits the largest coverage of publications with Mendeley readership counts (30,089 publications, 96.6%), only 30% of publications have exactly the same scores as reported by Mendeley (as based on our DOI-approach for querying the API). Hence, there is a strong disagreement in the readership scores (a total of 70% of publications with higher or lower scores) reported by Plum analytics and our Mendeley data collection approach.

When we compare the agreements between the rest of pairs of altmetric aggregators, Altmetric.com and Lagotto exhibit the strongest agreement in their Mendeley scores (42%), while Altmetric.com and Plum Analytics have a much lower agreement (26.9%). An even lower agreement (only 25.5% of the 30,086 overlapped publications) is found between Lagotto and Plum Analytics. An explanation for these discrepancies could be the fact that Plum Analytics merges the Mendeley counts for different identifiers across different versions of the same publication, while our approach was partly limited by the DOI-querying of the Mendeley API, which seems to be used by other aggregators, particularly Altmetric.com and Lagotto (see section 3.2.1).

Table 3. Analysis of (dis)agreement among aggregators in Mendeley readership counts.

	Mendeley.com (n=30,124)				**Altmetric.com (n=19,073)**				**Lagotto (n=30,117)**			
Readerships	OL.	E.	>	<	OL.	E.	>	<	OL.	E.	>	<
Altmetric. Com (n=19,073)	19,015	18,613	153	249								
%		97.9	0.8	1.3								
Lagotto (n= 30,117)	30,117	14,416	13,974	1,727	19,012	7,977	9,823	1,212				
%		47.9	46.4	5.7		42.0	51.7	6.4				
Plum Analytics (n=30,389)	30,089	9,027	10,531	10,531	19,057	5,120	6,974	6,963	30,086	7,676	7,815	14,595
%		30.0	35.0	35.0		26.9	36.6	36.5		25.5	26.0	48.5

OL.=Overlapped; E.=Equal

- *Tweets and tweeters counts*

When it comes to the analysis of Twitter data, Altmetric.com presents the highest number of (re)tweets as compared to other aggregators. Overall, we notice a lower agreement in the Twitter scores reported by all aggregators (Table 4) as compared to Mendeley. The largest set of overlapped publications is found between Altmetric.com and Lagotto, with 9,763 publications with Twitter activity recorded by both aggregators. Of this overlapped dataset, just 3,135 publications (32.1%) report exactly the same Twitter counts.

Table 4. Analysis of (dis)agreement among aggregators in Twitter counts (re)tweets, and distinct tweeters.

	Altmetric.com (n= 17,926)				**CrossRef ED (n=555)**				**Lagotto (n=9,973)**			
Tweets	OL.	E.	>	<	OL.	E.	>	<	OL.	E.	>	<
CrossRef ED (n= 555)	546	54	8 (5)*	484 (463) *								
%		9.9	1.5	88.6								
Lagotto (n= 9,973)	9,763	3,135	1,027	5,601	515	74	404	37				
%		32.1	10.5	57.4		14.4	78.4	7.2				
Plum Analytics (n=7,526)	7,356	2402	258	4,696	525	156	355	14	4,143	957	895	2,291
%		32.7	3.5	63.8		29.7	67.6	2.7		23.1	21.6	55.3

OL.=Overlapped; E.=Equal; *The values in the parentheses refer to number of tweeters - only available for Altmetric.com and CrossRef ED.

- *Facebook counts*

Regarding Facebook counts, the strongest agreement is between Plum Analytics and Lagotto with 1,130 publications (45.3%) with exactly the same scores (Table 5). The strongest discrepancies are found between Altmetric.com and Plum Analytics, with a total of 1,362 publications (74.9%) with higher scores in Plum Analytics than in Altmetric.com. Plum Analytics also has a total of 1,330 publications (53.3%) with higher scores than in Lagotto. Lagotto has higher Facebook counts in 770 publications (64.5%) with higher scores than in Altmetric.com (Table 5).

Table 5. Analysis of (dis)agreement among aggregators in Facebook counts.

	Altmetric.com (n=3,623)				Lagotto (n=2,497)			
Facebook counts	OL.	E.	>	<	OL.	E.	>	<
Lagotto (n=2,497)	1193	149	770	274				
%		12.5	64.5	23.0				
Plum Analytics (n=5,149)	1819	225	1362	232	2496	1130	1330	36
%		12.4	74.9	12.8		45.3	53.3	1.4

OL.=Overlapped; E.=Equal

- *Wikipedia counts*

In terms of Wikipedia citations, the strongest agreement is between Plum Analytics and Altmetric.com as 86.1% of the overlapped publications between them have exactly the same Wikipedia counts. Between Lagotto and Plum Analytics there are 65.6% of overlapped publications with exactly the same Wikipedia counts; while 62.2% of the overlapped publications between Lagotto and Altmetric.com have equal values of Wikipedia counts (Table 6). The lowest agreement is between the Wikipedia counts by CrossRef ED and all other aggregators, although CrossRef ED has systematically higher values of Wikipedia mentions than the other aggregators as showed in Table 2. Hence, the highest discrepancies between the Wikipedia counts reported by CrossRef ED and the others are mostly explained by the higher counts reported in this source.

Table 6. Analysis of (dis)agreement among aggregators in Wikipedia counts.

	Altmetric.com (n=639)				CrossRef ED (n=716)				Lagotto (n= 1,615)			
Wikipedia counts	OL.	E.	>	<	OL.	E.	>	<	OL.	E.	>	<
CrossRef ED (n=716)	464	74	367	23								
%		15.9	79.1	5.0								
Lagotto (n=1,615)	611	380	218	13	643	97	71	475				
%		62.2	35.7	2.1		15.1	11.0	73.9				
Plum Analytics (n=747)	612	527	42	43	518	97	21	400	697	457	8	232
%		86.1	6.9	7.0		18.7	4.1	77.2		65.6	1.1	33.3

OL.=Overlapped; E.=Equal

3.3.1.4. Correlation among metrics across aggregators

Previous sections have discovered important discrepancies in terms of coverage and counts among altmetric data aggregators. In this section we test the importance of the differences depicted so far using Pearson correlation analyses (Table 7). In order to reduce the effect of publications with zero values in all data aggregators, only publications with at least a non-zero score in any of the aggregators for each of the different social media platforms have been considered[14].

Mendeley readership counts

As shown in Table 7, a total of 30,433 (96.8%) publications have some readership scores from at least one data aggregator. The correlations among the different aggregators are relatively high (*r>.8*) in all cases. The lowest correlations are found between Altmetric.com and the other aggregators. This can be related to the lower coverage of Mendeley readership in Altmetric.com, which does not report Mendeley scores for many publications.

Table 7. Pearson Correlation analysis across different aggregators and their Mendeley readership counts.

N=30,433	**Altmetric.com**	**Lagotto**	**Mendeley**	**Plum Analytics**
Altmetric.com	1	.917	.918	.874
Lagotto		1	.998	.945
Mendeley			1	.946
Plum Analytics				1

Twitter counts[15]

Regarding Twitter counts, correlations vary between high and moderate between most pairs of aggregators. Altmetric.com has the highest correlation with Lagotto (*r=.9*) and Plum Analytics (*r=.7*) and together with Lagotto they have moderate correlations with CrossRef ED (between *r=.5* and *r=.6*). Also, there is a moderate correlation (*r=.5*) between Twitter users (tweeters) from Altmetric.com and CrossRef ED (Table 8).

Table 8. Pearson Correlation analysis across different aggregators and their Tweets and tweeters.

N=18,285	**Altmetric.com**		***CrossRef ED***		**Lagotto**	**Plum Analytics**
	Tweets	*tweeters*	*Tweets*	*tweeters*		
Altmetric.com						
Tweets	1	.979	.636	.602	.952	.762
tweeters		1	.593	.578	.955	.752
CrossRef ED						
Tweets			1	.983	.641	.516
tweeters				1	.622	.488
Lagotto					1	.728
Plum Analytics						1

[14] The idea is to minimize the effect of publications with zero in all the aggregators for the same indicator. Thus, we correlated only those publications with a value higher than zero in at least one aggregator, assuming a value of zero for the aggregators not reporting any value on the given metric.

[15] Twitter users (tweeters) refer to the number of users who have tweeted publications. This information is available for Altmetric.com and CrossRef Event Data.

Facebook counts

Correlations between Facebook counts are low for all pairs of aggregators. The highest correlation is between Plum Analytics and Lagotto (*r=.3*) and the weakest correlation is between Altmetric.com and these two platforms (around *r=.1*) (Table 9).

Table 9. Pearson Correlation analysis across different aggregators and their Facebook counts.

N=6,953	**Altmetric.com**	**Lagotto**	**Plum Analytics**
Altmetric.com	1	.112	.134
Lagotto		1	.397
Plum Analytics			1

Wikipedia counts

Correlations for Wikipedia counts range from high (*r=.8*) between Altmetric.com and Plum Analytics, moderate (*r=0.5*) between Altmetric.com and Lagotto, to weaker correlation (between *r=0.2* and *r=0.3*) among CrossRef ED and the other three aggregators (Table 10).

Table 10. Pearson Correlation analysis across different aggregators and their Wikipedia counts.

N=1,727	**Altmetric.com**	**CrossRef ED**	**Lagotto**	**Plum Analytics**
Altmetric.com	1	.380	.551	.867
CrossRef ED		1	.276	.388
Lagotto			1	.459
Plum Analytics				1

Based on the above results, Mendeley counts exhibit the highest correlations. Thus, Mendeley readership counts provided by all data aggregators are relatively consistent, although the coverage is limited in Altmetric.com. Regarding Twitter, correlations are moderate to high (with values ranging between *r=.4* and *r=.9*). Thus, tweets from Altmetric.com, Lagotto, and Plum Analytics are highly correlated among each other, while the lower correlations are found between CrossRef ED and the other aggregators. Similar levels of correlation between Mendeley readership and tweets across similar altmetric data aggregators have been observed in a previous study for publications from two journals in the Library and Information science field (Bar-Ilan and Halevi, 2017). Regarding Wikipedia counts, Plum Analytics and Altmetric.com are strongly correlated (*r=.8*), which is also related to the stronger agreement between these two aggregators in Wikipedia counts. Similar correlations for Wikipedia mentions between Altmetric.com and Plum Analytics have been observed by a recent study for a random sample of 5,000 Web of Science publications from the year 2015 (Meschede & Siebenlist, 2018). However, the correlations for Wikipedia counts among the other combinations of aggregators are in general rather weak or just moderate, ranging between *r=.2* (for Lagotto and CrossRef ED) and *r=.5* (for Lagotto and Altmetric.com). Facebook counts is the source with the lowest correlations overall. Although Facebook counts from Lagotto and Plum Analytics exhibit the highest correlation compared to all other aggregators, the correlation is just of *r=.3*. The correlations of Facebook counts with Altmetric.com are in all cases very weak.

3.3.2. Reasons for differences found across altmetric data aggregators

Although the metrics have been collected at the same time for the same dataset, the results presented above demonstrate that there are relevant differences in the publications covered by the different altmetric aggregators, as well as in the data collected and reported by them. An overview of the main methods of collecting, processing, and reporting altmetric data from the altmetric aggregators considered in this study is presented in Table 11. The information in this table is obtained from the websites of the different aggregators, as well as from the information they have reported in the NISO altmetrics code of conduct (National Information Standards Organization, 2016). Based on Table 8, in this section, we reflect over the possible reasons for these differences, trying to provide more insights based on additional observations extracted from the data collected. The focus is more on the discussion of the effects of methodological choices than on the benchmark of altmetric aggregators. It is of course very difficult to depict all the underlying reasons for the differences found due to the lack of information on how each aggregator specifically queries and processes the original data sources. However, we argue that most of the data issues identified can be conceptually related to the following four major groups of methodological choices: 1) *data collection choices*; 2) *data aggregation and reporting choices*; 3) *updating choices;* and 4*) other technical choices*.

Table 11. Overview of the main methods of collecting, tracking, and updating metrics across different altmetric data aggregators – as reported by the data aggregators.

Social media sources	Aggregators	Data collection approaches	Data aggregation and reporting approaches			Data updating approaches
		API use	*Objects and identifiers*	*Aggregated metrics*	*Raw data & provenance*[16]	*Updates*
Mendeley Readership	**Altmetric.com**	Mendeley API.	Tracks, orderly, scholarly objects with DOI, PMID, ArXiv ID and stops the process if any result is found by any of the identifiers.	Aggregated individual user readership counts.	Raw data on readership by academic types, countries, and disciplines is recorded.	Daily updates.
	Plum Analytics	Is part of Elsevier and does not directly use the Mendeley API.	Tracks any identifiers (DOIs, PMIDs, etc.)		Raw data is not provided.	
	Lagotto	Mendeley API.	Tracks DOIs.	Aggregated individual user and group readership counts.	Raw data is not provided.	
Twitter	**Altmetric.com**	Twitter GNIP API.	Tracks a range of different identifiers (URLs, DOIs, PMIDs, ArXiv ids , SSRN IDs, ADS IDs, Amazon URLs, and ISBNs)	Aggregated count of distinct tweeters. Aggregated counts of (re)tweets provided in the Bookmarklet.	Raw data from Twitter (tweets, retweets, tweeters, followers, etc.) is available in the JSON files through the Altmetric.com API	Real-time update.
	Plum Analytics		Tracks a range of different identifiers (URLs, DOIs, PMIDs, PMCID, ArXiv IDs, ISBNs, etc.[17])	Aggregated counts of (re) tweets across multiple versions of the same output.	Raw data is not provided.	Real-time update.
	CrossRef ED	Twitter GNIP Power Track API.	Tracks DOIs and article landing page URLs. No other identifiers.	Only raw data is provided.	Raw data from Twitter (tweets, retweets, tweeters, followers, etc.)	Real-time update.
	Lagotto	Twitter Search API with rate limit of 1,800 requests per hour.	Tracks DOIs and journal landing page URLs.	Aggregated counts of (re)tweets.	Raw data is not provided.	No information.

[16] As obtained through our querying method. See also Methods section.
[17] https://plumanalytics.com/identifiers-types-research-output/

Social media sources	Aggregators	Data collection approaches	Data aggregation and reporting approaches			Data updating approaches
		API use	*Objects and identifiers*	*Aggregated metrics*	*Raw data & provenance*[16]	*Updates*
Facebook	**Altmetric.com**	Facebook Graph API.	Same as for Twitter.	Aggregated counts of public Facebook posts.	Raw data is not provided.	No information.
	Plum Analytics		Same as for Twitter.	Combined counts of all public and private Facebook likes, shares, and comments.		Daily update.
	Lagotto		Tracks journal landing page URLs.		Raw data is not provided.	No information.
Wikipedia	**Altmetric.com**	Wikipedia API.	Tracks all Wikipedia edits searching for links to scholarly domains, and also clearly labeled identifiers (DOIs and PMIDs).	Count of Wikipedia mentions in the references of English pages[18].	Raw data is available.	Real-time update.
	Plum Analytics	Retrieves Wikipedia mentions by mining search engine results, watching Wikipedia pages for citation changes, and mining the full text of all Wikipedia pages.	Tracks only URLs.	Count of Wikipedia mentions in the references of English pages[19].	Raw data is not provided.	Daily update.
	Lagotto	MediaWiki API.	Tracks DOIs and URLs.	Aggregated score of mentions in the Wikipedia pages and files.	Wikipedia mentions in the references of the 25 most popular languages of Wikipedia pages [20]	No information

[18] We have used 2016 version of Altmetric.com data for this study. From 2017 Altmetric.com announced the tracking of some non-English (Swedish and Finnish) pages in Wikipedia https://help.altmetric.com/support/solutions/articles/6000060980-how-does-altmetric-track-mentions-on-wikipedia-.
[19] Plum Analytics announced the tracking of Spanish and Portuguese language Wikipedia entries https://plumanalytics.com/spanish-portuguese-wikipedia-references-now-plumx-metrics/
[20] http://www.lagotto.io/docs/wikipedia/

Social media sources	Aggregators	Data collection approaches	Data aggregation and reporting approaches			Data updating approaches
		API use	*Objects and identifiers*	*Aggregated metrics*	*Raw data & provenance*[16]	*Updates*
					and Wikimedia commons (repository of media files).	
	CrossRef ED	Wikipedia MediaWiki Event Streams and the MediaWiki APIs.	Only tracks DOIs and article landing page URLs (not any other identifiers).	Count is not provided.	Raw data on Wikipedia mentions in the references of both old and new versions (edits) of English and non-English pages.	Real-time update.

3.3.2.1. Data collection choices

Metrics depend largely on the way each aggregator collects the data from the related social media sources. This can be done directly from the original social media platform, or indirectly through a third-party vendor, bot, or agent (Fenner, 2013). In this case, the use of different APIs can partly explain the differences in the values of metrics reported by the altmetric aggregations. Additionally, the focus on different identifiers, URLs, landing pages, or scholarly objects can also provide different results.

Mendeley

Except Plum Analytics, Altmetric.com and Lagotto use the same Mendeley REST API. However, the specific way each aggregator queries the API using specific or multiple identifiers can have an effect on the final reported readership counts. For instance, Altmetric.com only queries Mendeley readership when other altmetric event has been reported for the publication (Peters et al., 2017; Robinson-Garcia, Torres-Salinas, Zahedi & Costas, 2014). This largely explains the strong discrepancies between this aggregator and the others in the reporting of Mendeley readership. Also, Altmetric.com queries Mendeley first using the DOI of the publication, and if no record is found then the PMID, and subsequently the ArXiv id are used. It stops the API when one of the identifiers has provided a match. Lagotto queries Mendeley API using only DOIs. In contrast, Plum Analytics does not query Mendeley API since it gets the data from Mendeley as it is part of Elsevier. Plum Analytics cross-reference all identifiers for a given object against data that they get from Mendeley and provides aggregated readership counts for all versions of the same object.

Hence, these different approaches can explain why Altmetric.com has a much lower coverage of publications with Mendeley readership, although its agreement with our DOI-approach method for Mendeley is strong. In contrast, Plum Analytics exhibits higher values and coverage, but has lower agreement with other aggregators.

Twitter

The collection of tweets using the Twitter Search API differs significantly from using the Streaming APIs[21] (see Appendix 1, Text 1 for an explanation of the differences between them). The use of Twitter APIs provided by third-party companies such as GNIP could also influences the metrics provided by the different altmetric aggregators. For example, GNIP offers access to real-time data with the possibility of filtering the searches based on keywords, geo-locations, etc. This in turn requires payment and depends on GNIP's sale agreement with the social media source (Twitter). Moreover, the frequency and the type of query used to use the Twitter API by the altmetric aggregators, also influence the metrics provided by them (Batrinca and Treleaven, 2015).

[21] https://dev.twitter.com/rest/public/search

Most of the analyzed altmetric aggregators use GNIP for collecting tweet mentions, except Lagotto that uses the Twitter search API with a rate limit of 1,800 queries per hour[22]. However, altmetric aggregators also query differently the Twitter API (see Appendix 1, Text 2 for the methodological descriptions of Altmetric.com, CrossRef ED, Lagotto, and Plum Analytics). Unfortunately, there is no direct way to explore how each aggregator exactly uses these third party APIs and how their algorithms collect and process the metrics. Based on this, we can just indirectly conclude that the choice of the Twitter API with rate limit could explain both why Lagotto reports substantially lower scores in the overlapped publications with both Altmetric.com and Plum Analytics. Additionally, the focus on mentions to just DOIs and journal landing page URLs for the registered contents by CrossRef ED can explain the lower Twitter coverage and counts provided by this aggregators in contrast with other aggregators (Altmetric.com and Plum Analytics[23]) that typically also use other identifiers (e.g., PMID, ArXiv id, etc.) to track the mentions to the papers. Also the recent start of CrossRef ED may imply that they have started to collect tweets from their inception but not retrospectively, thus explaining why the other altmetric aggregators exhibit so much higher Twitter counts. Furthermore, how each aggregator accommodates the Twitter compliance guidelines also influences the data reported for each tweet. For instance, if a user deletes a tweet, changes its sharing options from public to protected or withheld, or has their account deleted or suspended, the tweet will be immediately removed and will not be displayed by Plum Analytics[24] (which would imply a decrease in the number of tweet counts for the publication). In contrast, Altmetric.com reports the deleted tweets but doesn't display the tweets that are no longer public[25] (which implies that the count won't decrease but wouldn't be possible to fully recreate it).

Facebook counts

All the three aggregators that collect Facebook mentions (Altmetric.com, Lagotto, and Plum Analytics) use the same Facebook Graph API. However, with respect to the choice of identifiers, Lagotto uses journal landing page URLs while Plum Analytics and Altmetric.com use any identifiers for collecting Facebook mentions (same approach as for Twitter). Hence, the use of additional identifiers by Plum Analytics explains the higher Facebook counts than Lagotto. Furthermore, the choice of aggregators in collecting public or private scores also results in variations in the metrics offered by them. In this case, Lagotto and Plum Analytics collect both Facebook public and private posts while Altmetric.com collects only public wall posts. Additionally Plum Analytics and Lagotto also count other Facebook events such as likes, shares, and comments. These choices explain the substantially lower Facebook coverage and values reported by Altmetric.com compared to Lagotto and Plum Analytics.

[22] http://www.lagotto.io/docs/twitter_search/
[23] https://plumanalytics.com/niso-altmetrics-working-group-on-data-quality/
[24] http://support.gnip.com/apis/consuming_compliance_data.html
[25] https://www.altmetric.com/details/3946203/twitter

Wikipedia counts

With respect to Wikipedia mentions, the choice of API also differs across aggregators. For example, Altmetric.com uses the Wikipedia API, Lagotto uses the MediaWiki API, CrossRef ED uses the MediaWiki Event Streams and the MediaWiki API. In contrast, Plum Analytics retrieves Wikipedia mentions by combining different methods of mining search engine results, the full text of all Wikipedia pages, and looking into Wikipedia pages for citation changes. Clearly, all these approaches could contribute to explain the differences in the Wikipedia counts reported by these aggregators. Moreover, other reasons for the observed differences relate to the choice of aggregations are explained in the next section.

3.3.2.2. Data aggregation and reporting choices

Metrics can largely be influenced by how scores for different versions of the same object with possible different identifiers[26] (DOI, PMID, ArXiv ID, URL) or versions (e.g., the ArXiv versions, the published version) are aggregated as well as how different events coming from the same social media platform are combined and reported. Moreover, aggregations could be based on different languages, edits, types, and scholarly objects.

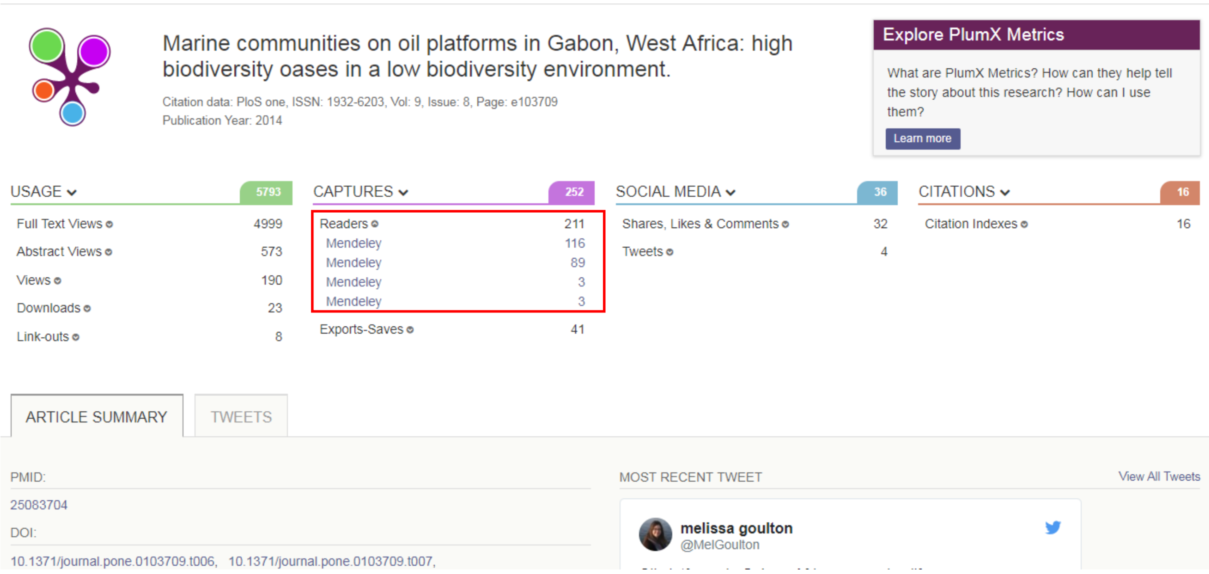
3.3.2.2.1. Aggregations of different identifiers and versions of the same publication

The metrics calculated for a publication may also depend on the quality of the metadata for which social media events are collected as well as on the existence of duplicate records of the same publication. Differences in metrics may arise when altmetric data aggregators handle differently these duplicates and merge (or not) different scores coming from multiple identifiers or versions of the same object.

Both Lagotto and Plum Analytics provide the largest coverage of publications with at least one reader and the highest counts of Mendeley readership. A potential explanation for these higher counts can be the merging of counts from different identifiers (e.g., DOI and PMIDs) for the same publication. This seems to be the case for Plum Analytics (see Appendix 1, Text 3). The problem is that the merging of counts from different identifiers can also imply some degree of error. For example, wrong linkages between identifiers happen and may create over or under-merging of records. Considering that Mendeley is a user-driven database, users may create wrong linkages between PMIDs, DOIs, etc. and therefore this sometimes causes errors in the assignment of readership to publications. Figure 1 shows how the linkages to both wrong DOIs and PMIDs in Mendeley leads to higher Mendeley readership reported by Plum Analytics. Plum Analytics aggregates readership (211 reader counts instead of 89 reader counts from Mendeley) for the paper with both correct and incorrect DOIs and PMIDs in Mendeley[27] (See Appendix 1, Figure 1) and thus, incorrectly reports 122 higher reader counts.

[26] A journal article on a publisher platform with given a DOI could be indexed by PubMed or an institutional repository with different unique identifiers (PubMed or ArXiv IDs).

[27] This paper has the PMDI=25083704, and the following Mendeley records wrongly contain also the same PMID: (https://www.mendeley.com/research-papers/m%C3%A9todos-para-medir-la-biodiversidad-23/



https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0103709&display-tab=summary-content#READER_COUNT

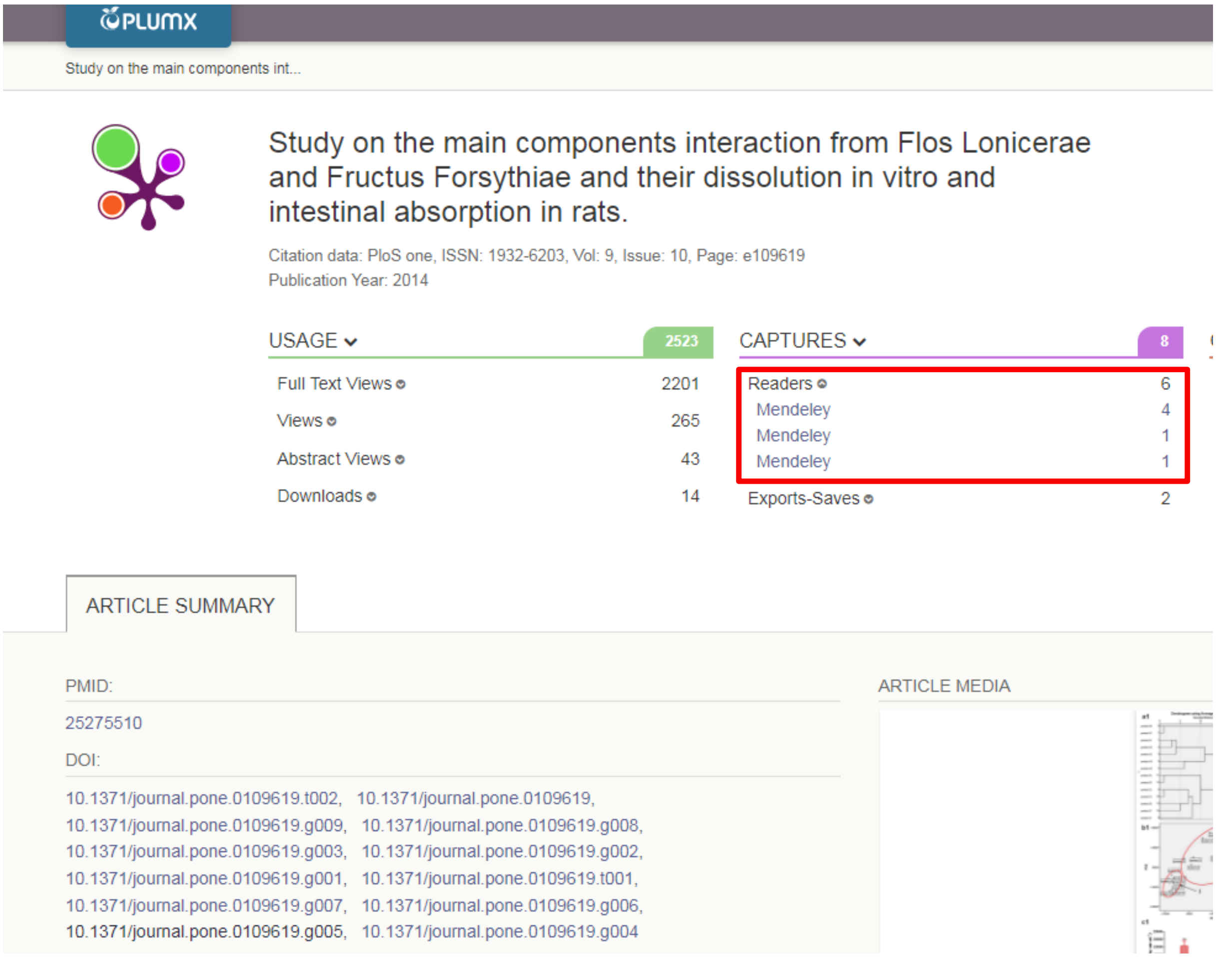
https://www.mendeley.com/research-papers/marine-communities-oil-platforms-gabon-west-africa-high-biodiversity-oases-low-biodiversity-environm/

Figure 1. Examples of different readership counts across different altmetric aggregators: Plum Analytics vs. Mendeley (accessed on 15 December 2017).

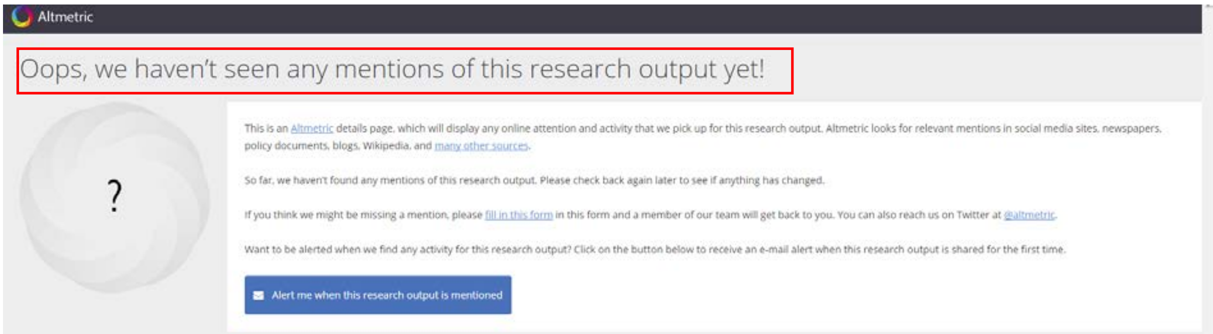
---

https://www.mendeley.com/research-papers/systematic-distribution-orihelia-anticlava-molin-1858-nematoda-onchocercidae-dasypodids-south-americ/
https://www.mendeley.com/research-papers/m%C3%A9todos-para-medir-la-biodiversidad-43/)

Also, when the same object appears with different records in Mendeley, for example one with a DOI and another one with a PMID, Plum Analytics aggregates all the readership counts across all the different versions. Figure 2 provides an example of how Plum Analytics aggregates Mendeley readership counts (6 counts in total across 2 versions) for all the multiple versions (duplicate records) in Mendeley of the same object. This example illustrates how wrong linkages to PMID in Mendeley affects the total readership counts reported. The readership counts across three different versions of the same object is reported in Figure 2. The record with 4 counts is recorded with both DOI and PMID while the other two records with 1 count each have only DOI or only PMID in Mendeley (See Appendix 1, Figure 2). However, the record with PMID in Mendeley has a wrong linkage to PMID and this leads to incorrectly reporting one extra readership count by Plum Analytics. Lagotto finds the record with only 1 readership for this object and fails to report the record with 4 readerships in Mendeley. Altmetric.com fails to find any readership for the same object since it is not mentioned in any other sources it tracks.



https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0109619#READER_COUNT



https://www.altmetric.com/details/doi/10.1371/journal.pone.0109619

http://alm.plos.org/works?q=10.1371%2Fjournal.pone.0109619&source_id=mendeley

Figure 2. Examples of different readership counts across different altmetric aggregators: Plum Analytics, Altmetric.com, and Lagotto vs. Mendeley (accessed on 15 December 2017).

3.3.2.2.2. Aggregations of different events from the same social media platform

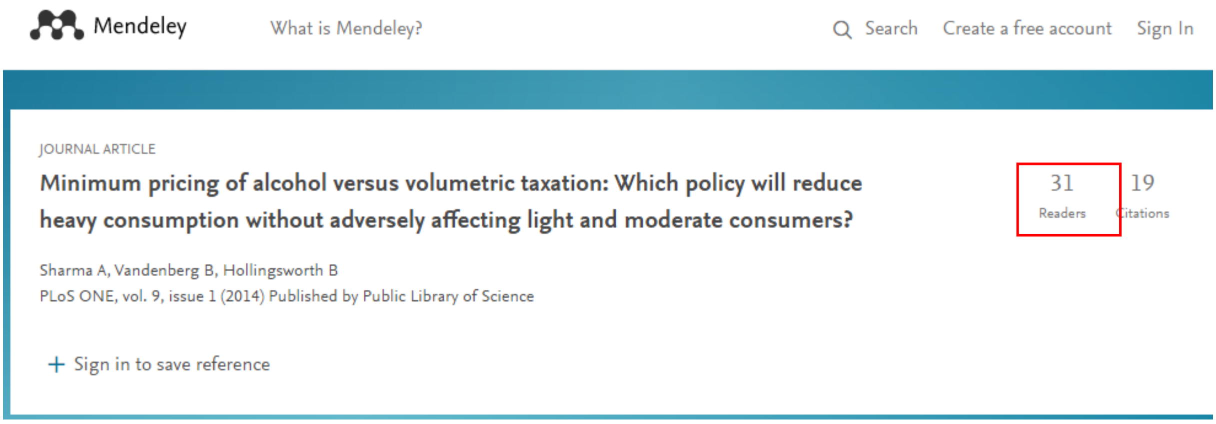
Aggregators may use the same API to query some data sources such as Twitter, Facebook, or Mendeley (i.e., Twitter GNIP API[28],Facebook Graph API[29], or Mendeley API); however, they could differ in the ways they combine different forms of scores from the same social media platform.

- *Mendeley*

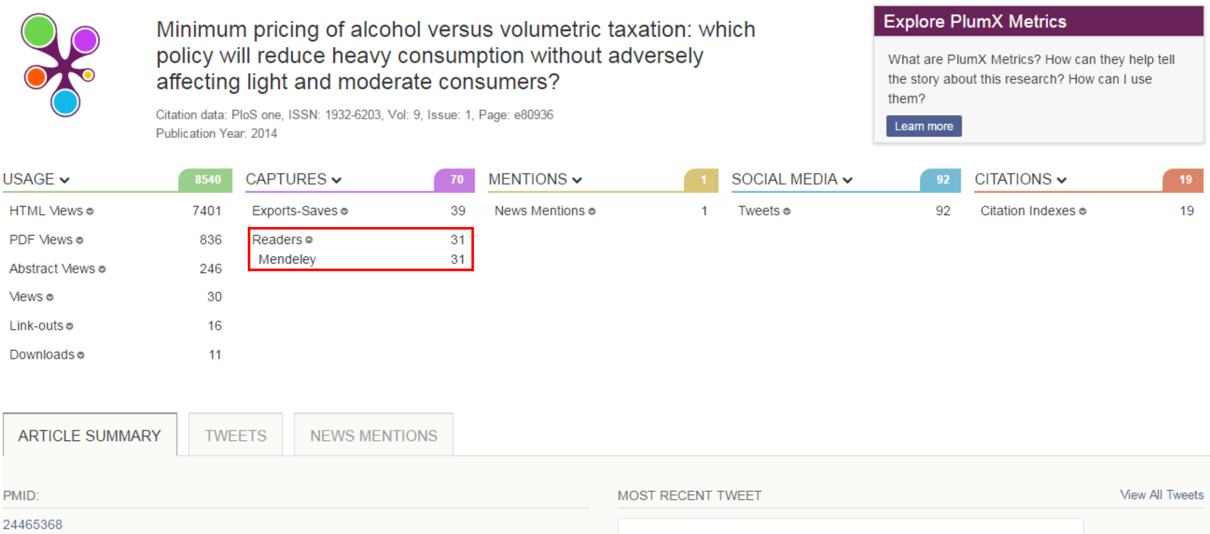
The higher readership values reported by Lagotto in comparison to the other aggregators can be explained by Lagotto's choice of reporting combined readership values of individual user and group counts. As an example in Figure 3, a publication with 31 readership according to Mendeley is presented, for which Lagotto reports 31 readers, plus three "group count" readers, making a total of 34; while Altmetric.com and Plum Analytics both report 31 readers each.

[28] http://support.gnip.com/apis/

[29] https://developers.facebook.com/docs/graph-api/

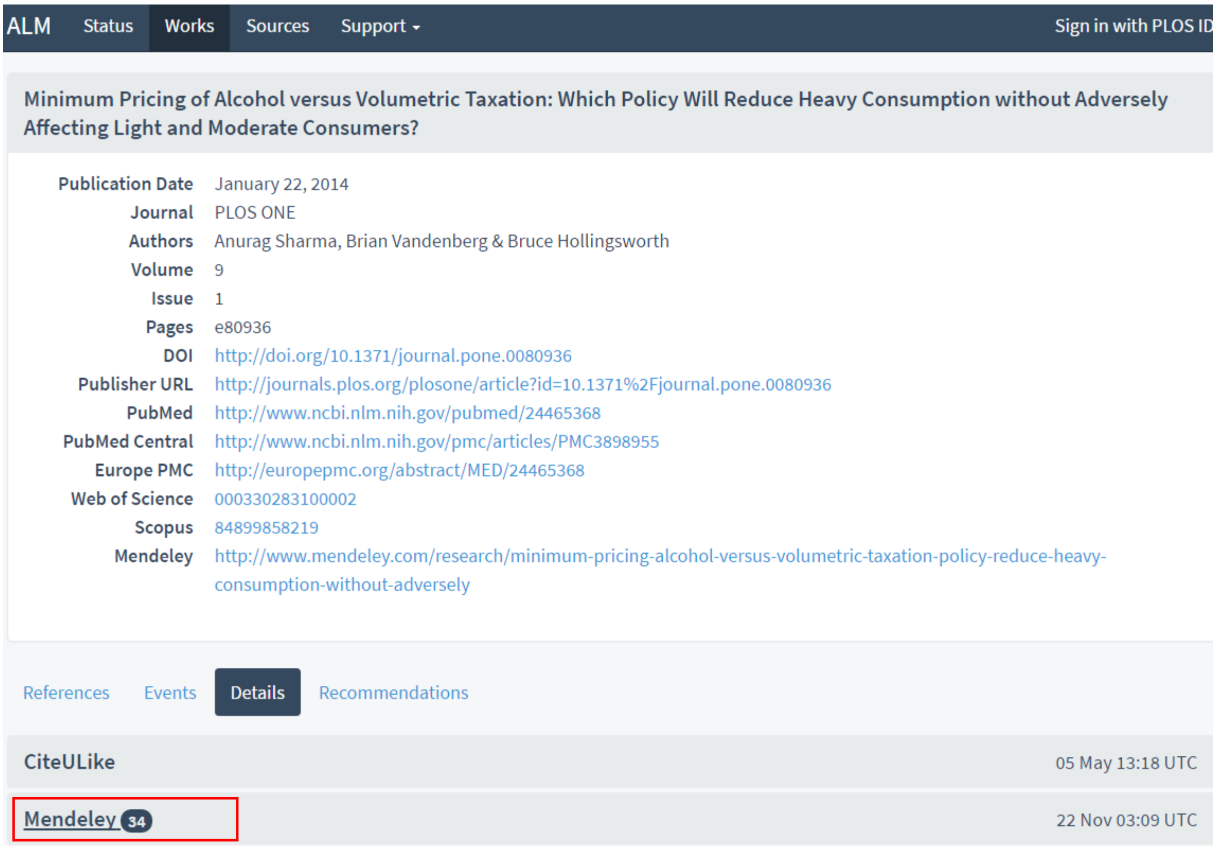


https://www.mendeley.com/research-papers/minimum-pricing-alcohol-versus-volumetric-taxation-policy-reduce-heavy-consumption-without-adversely/



https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0080936&display-tab=summary-content#READER_COUNT

https://www.altmetric.com/details/2065475



http://alm.plos.org/works/doi.org/10.1371/journal.pone.0080936

Figure 3. Examples of different Mendeley readership counts across different altmetric aggregators: Mendeley, Plum Analytics, Altmetric.com, and Lagotto are presented orderly (accessed on 29 November 2017).

- *Facebook*

Figure 4 illustrates an example of large differences in Facebook counts. Lagotto reports a total 1,023 combined Facebook score of all activities (posts, shares, likes, comments), while Altmetric.com reports 264 Facebook public post counts (excluding likes, individual timeline, and private posts). Plum Analytics exhibits the largest number of Facebook counts (27,296) including the sum of all public and private Facebook likes, shares, and comments across the multiple identifiers that it tracks for the same publication. Unfortunately the lack of access to the raw data does not allow exploring further the reasons for such large difference between Lagotto and Plum Analytics.

https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0088278&display-tab=summary-content&tweet-reference-id=all-tweets-tab#PLUS_ONE_COUNT,TWEET_COUNT,FACEBOOK_COUNT

http://alm.plos.org/works?q=10.1371%2Fjournal.pone.0088278&source_id=facebook

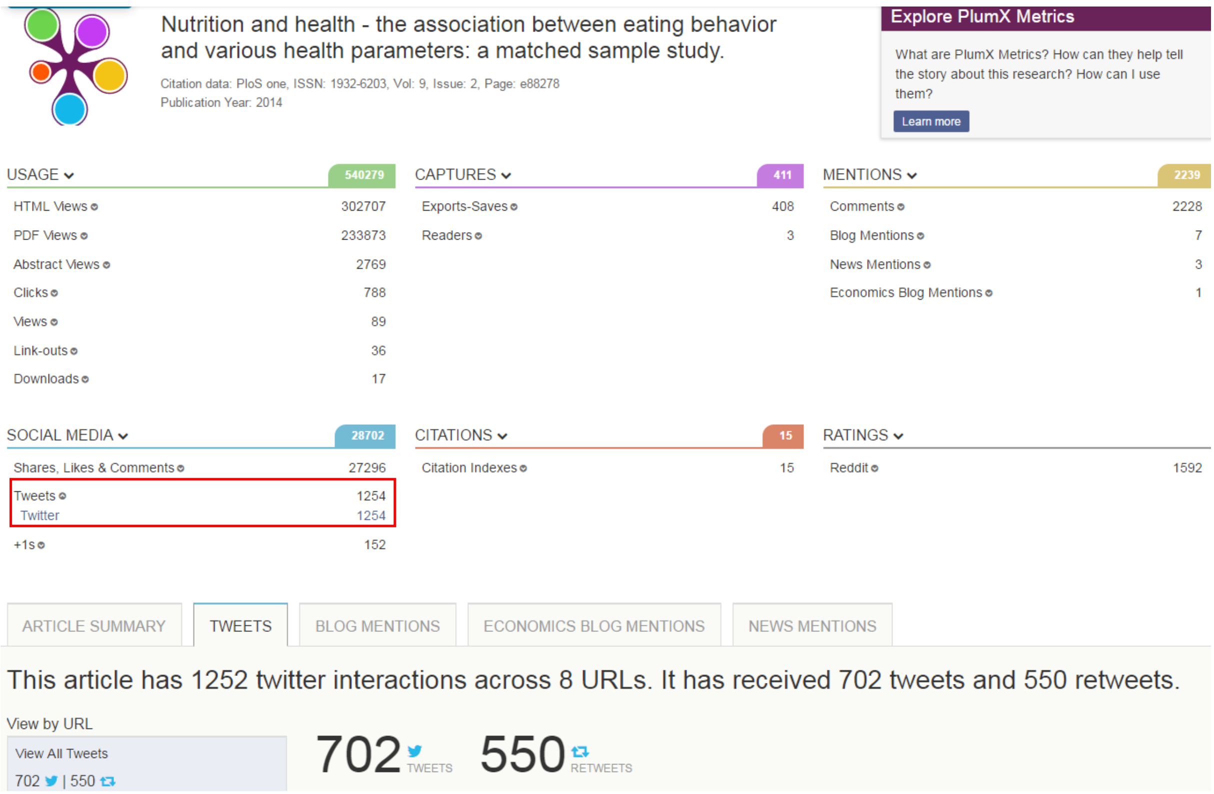
https://www.altmetric.com/details/2108919/facebook

Figure 4. Examples of different Facebook counts across different altmetric aggregators: Plum Analytics, Lagotto, and Altmetirc.com are presented orderly (accessed on 29 November 2017).

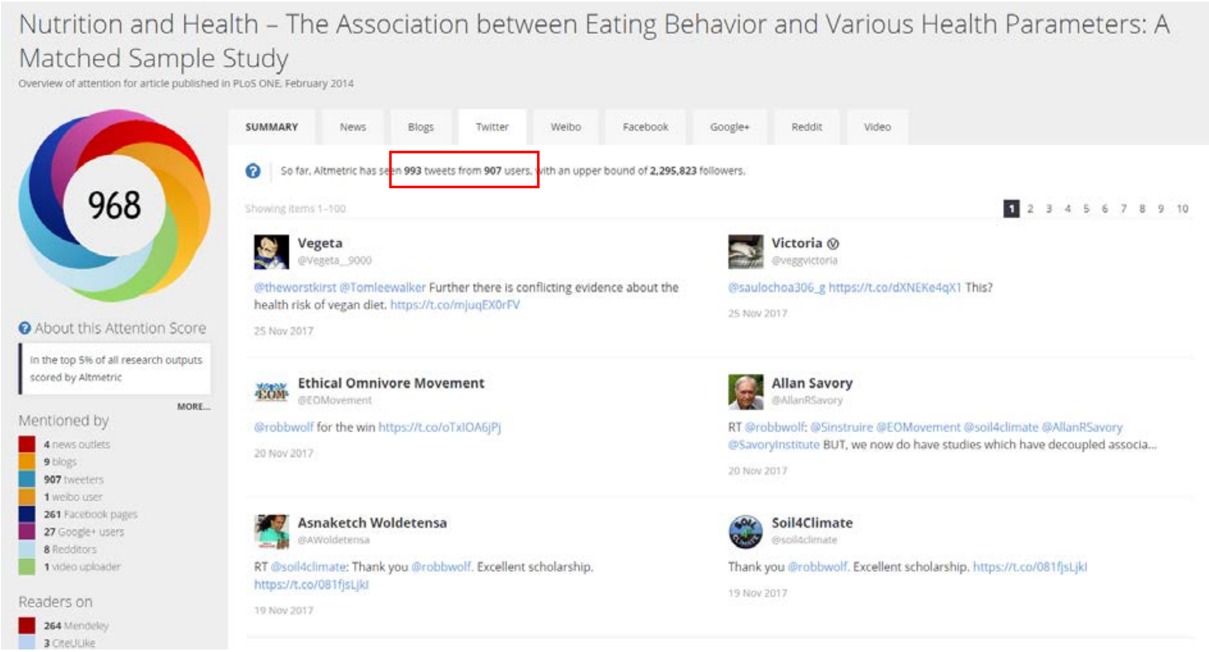
- *Twitter*

Most of the altmetric aggregators report different indicators on Twitter events. For example, in case of Altmetric.com and Lagotto, both tweets and retweets are combined for each publication (although the indicator promoted by Altmetric.com is the number of distinct tweeters, both from tweets and retweets). Plum Analytics reports separately the number of tweets and retweets that mention the object, but it combines both counts in a final Twitter score in the main summary. Plum Analytics also reports combined scores of (re)tweets for the different versions of the same object. CrossRef ED also reports all data of tweets and retweets. The higher value of tweets reported by Altmetric.com compared to other aggregators could be explained by the fact that "Altmetric.com collates the scores for the different version of the same research output" (see Appendix 1, Text 5). However, it is not clear why CrossRef ED provides the least tweets than other aggregators. A potential explanation is that as publications from 2014 analyzed in this study while CrossRef ED has started in April 2017, tweets between 2014 and 2017 could be missing.

Figure 5 illustrates an example of the Twitter counts from these aggregators for the same publication. Overall, Plum Analytics reports 1,254 (702 tweets and 550 retweets) across the 8 different URLs pointing to the same publication from different databases such as PubMed, PMC, and PloS ONE. Hence the tweet value is higher than the values reported by Altmetric.com (907) and Lagotto (941) while CrossRef ED reports no tweets for this publication.



https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0088278&display-tab=artifact-tweets&tweet-reference-id=all-tweets-tab#TWEET_COUNT



https://www.altmetric.com/details/2108919/twitter

http://alm.plos.org/works/doi.org/10.1371/journal.pone.0088278

Figure 5. Examples of different tweets (tweeters) across different altmetric aggregators: Plum Analytics, Altmetric.com, and Lagotto are presented orderly (accessed on 29 November 2017).

3.3.2.2.3. Aggregation based on languages, document types, scholarly objects, and edits

The choice of aggregators in aggregating scores for particular data sources, document types, languages, or scholarly objects could also influence the metrics provided. For instance, discrepancies in the value of Wikipedia mentions can be explained by the different approaches in aggregating DOIs mentioned across Wikipedia pages in different languages as well as non-encyclopedia pages (such as user, talk, and Meta-wiki pages[30], media and files). Also, consideration of the edits of each Wikipedia page as separate events influences the counts. For instance, the Wikipedia mentions count reported by Lagotto is a combined score of the number of references to papers, material files, images, etc.[31] [10] from the 25 most popular Wikipedia sites[32]. In contrast, Altmetric.com[33] reports Wikipedia mentions of scholarly outputs collected from the reference sections of English, and from 2017 onwards Finnish, and Swedish languages Wikipedia entries. Altmetric.com doesn't track non-encyclopedic pages[34]. For Altmetric.com also every Wikipedia mention has to have an author, a timestamp, and valid citations such as title, PubMed ID, or DOI to be tracked. Wikipedia mentions reported by CrossRef ED includes 'edits of articles in Wikipedia' and combines mentions from all old and new versions[35], thus every edit of the paper is considered separately. This explains the higher Wikipedia mentions recorded by CrossRef ED. However, the approach taken by Plum Analytics for tracking Wikipedia mentions is different from the others. Plum Analytics uses a combination of data mining both in the search engine results and in open source repository platforms (such as Dspace[36]) and watches citation changes[37] in the Wikipedia pages. It mines full text of Wikipedia English pages (from March 2018 onwards Spanish and Portuguese are also tracked) and looks for any links to the object[38] (DOIs, PMIDs, URLs). Also, Plum Analytics tracks scholarly objects (thesis, book chapters, books, and technical reports) other than articles. The example in Figure 6 shows that Lagotto reports 10 Wikipedia mentions while the other aggregators do not report any for the same publication. This is because Lagotto, besides papers, also tracks files that contains links to papers, while the other aggregators don't do that.

---

[30] A free software open source wiki: https://www.mediawiki.org/wiki/MediaWiki

[31] http://alm.plos.org/works/doi.org/10.1371/journal.pone.0083259
https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0083259&display-tab=artifact-wikipedia#LINK_COUNT
https://www.altmetric.com/details/2029983/wikipedia

[32] http://www.lagotto.io/docs/wikipedia/

[33] https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track-

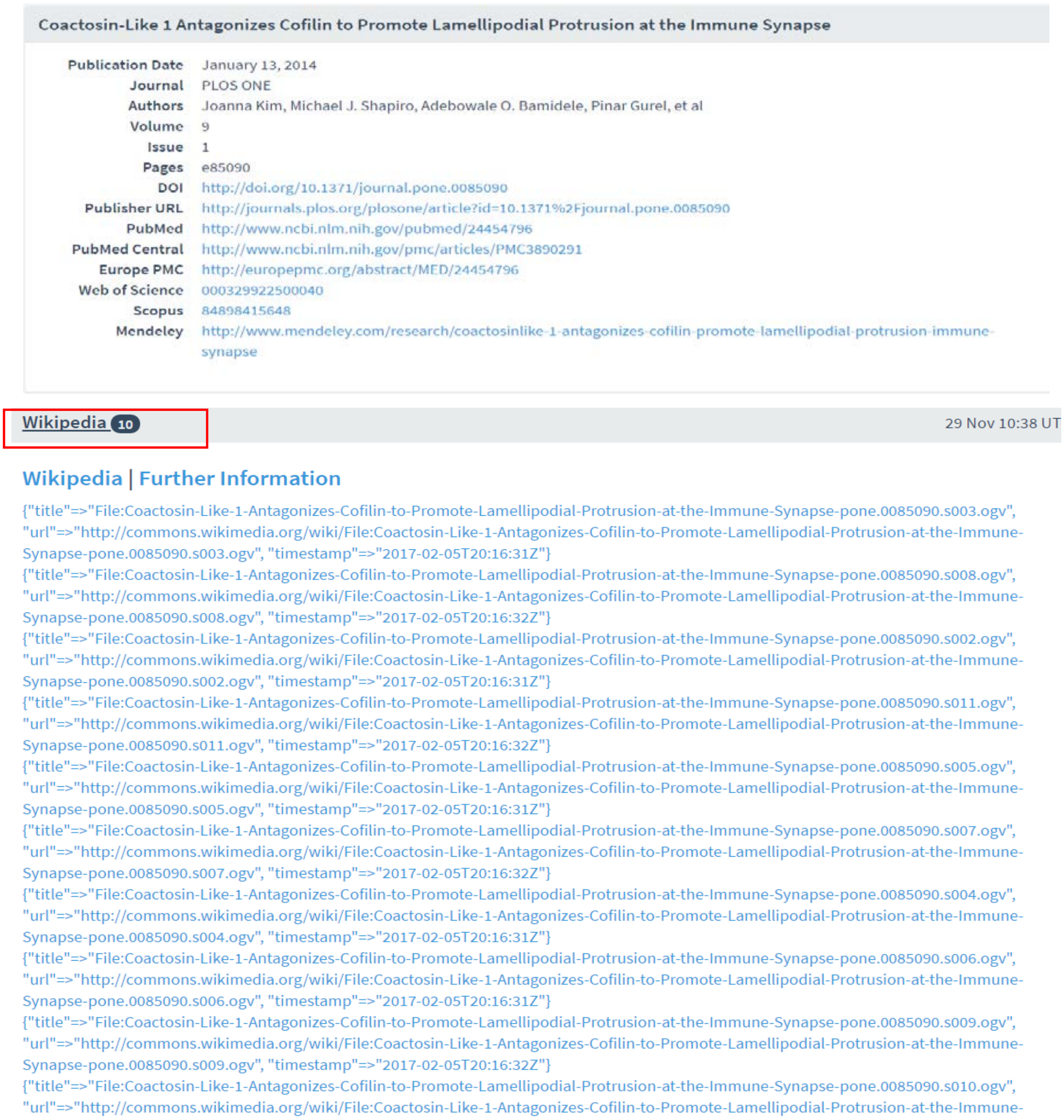
[34] https://help.altmetric.com/support/solutions/articles/6000060980-how-does-altmetric-track-mentions-on-wikipedia-

[35] https://en.wikipedia.org/w/index.php?title=Finger&oldid=783699149
https://en.wikipedia.org/w/index.php?title=Finger&oldid=783698538

[36] www.dspace.org/introducing

[37] http://plumanalytics.com/wikipedia-altmetrics-calculating-mention-metrics/
http://plumanalytics.com/wikipedia-altmetrics-context-completeness-and-what-mention-metrics-mean/

[38] https://plumanalytics.com/wikipedia-altmetrics-calculating-mention-metrics/


Coactosin-Like 1 Antagonizes Cofilin to Promote Lamellipodial Protrusion at the Immune Synapse

Publication Date	January 13, 2014
Journal	PLOS ONE
Authors	Joanna Kim, Michael J. Shapiro, Adebowale O. Bamidele, Pinar Gurel, et al
Volume	9
Issue	1
Pages	e85090
DOI	http://doi.org/10.1371/journal.pone.0085090
Publisher URL	http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0085090
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/24454796
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3890291
Europe PMC	http://europepmc.org/abstract/MED/24454796
Web of Science	000329922500040
Scopus	84898415648
Mendeley	http://www.mendeley.com/research/coactosinlike-1-antagonizes-cofilin-promote-lamellipodial-protrusion-immune-synapse

Wikipedia 10 29 Nov 10:38 UT

Wikipedia | Further Information

{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s003.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s003.ogv", "timestamp"=>"2017-02-05T20:16:31Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s008.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s008.ogv", "timestamp"=>"2017-02-05T20:16:32Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s002.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s002.ogv", "timestamp"=>"2017-02-05T20:16:31Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s011.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s011.ogv", "timestamp"=>"2017-02-05T20:16:32Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s005.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s005.ogv", "timestamp"=>"2017-02-05T20:16:31Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s007.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s007.ogv", "timestamp"=>"2017-02-05T20:16:32Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s004.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s004.ogv", "timestamp"=>"2017-02-05T20:16:31Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s006.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s006.ogv", "timestamp"=>"2017-02-05T20:16:31Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s009.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s009.ogv", "timestamp"=>"2017-02-05T20:16:32Z"}
{"title"=>"File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-Synapse-pone.0085090.s010.ogv", "url"=>"http://commons.wikimedia.org/wiki/File:Coactosin-Like-1-Antagonizes-Cofilin-to-Promote-Lamellipodial-Protrusion-at-the-Immune-

http://alm.plos.org/works/http:%2F%2Fdoi.org%2F10.1371%2Fjournal.pone.0085090?source_id=wikipedia

Figure 6. Examples of Wikipedia counts for an object reported by Lagotto (accessed on 29 November 2017).

The example depicted in Figure 7 shows a paper for which Lagotto reports 8 mentions (three in English language Wikipedia pages + 5 files). Altmetric.com reports 3 mentions (actually 4 citations found on 3 English language pages, but Altmetric.com only counts the number of distinct Wikipedia pages citing the publication). Plum Analytics also reports 3 mentions (in 3 English language pages) while CrossRef ED records 315 mentions (English and Macedonian language pages + edits made at different times) (See Appendix 1, Text 6).

Standing at the Gateway to Europe - The Genetic Structure of Western Balkan Populations Based on Autosomal and Haploid Markers

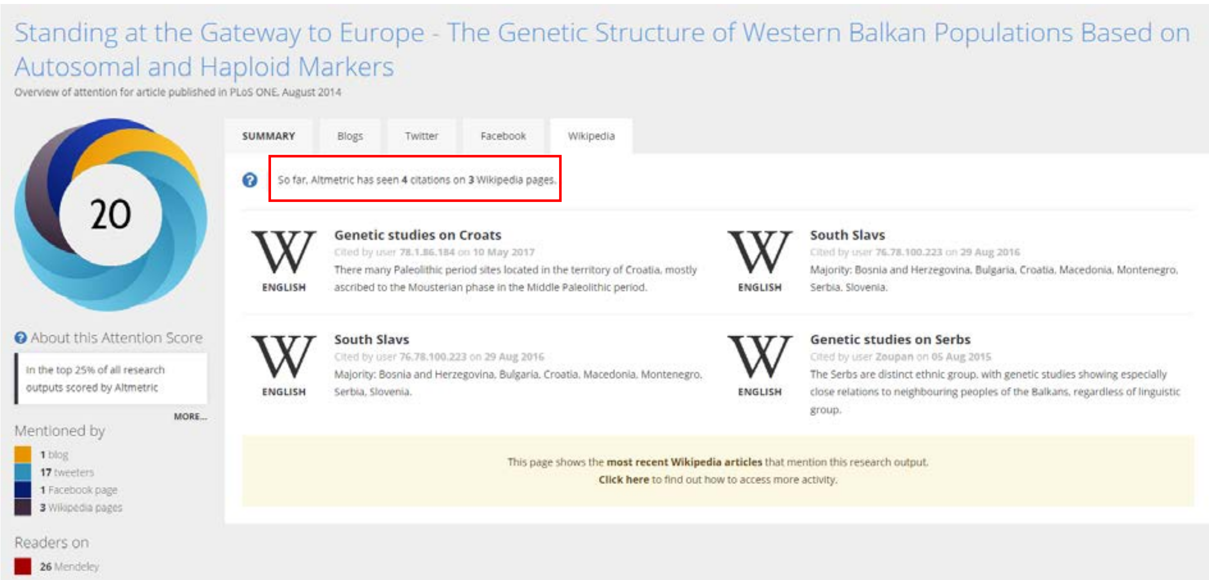
Publication Date	August 22, 2014
Journal	PLOS ONE
Authors	Lejla Kovacevic, Kristiina Tambets, Anne Mai Ilumäe, Alena Kushniarevich, et al
Volume	9
Issue	8
Pages	e105090
DOI	http://doi.org/10.1371/journal.pone.0105090
Publisher URL	http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0105090
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/25148043
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141785
Europe PMC	http://europepmc.org/abstract/MED/25148043
Web of Science	000341230600031
Scopus	84925865749
Mendeley	http://www.mendeley.com/research/standing-gateway-europe-genetic-structure-western-balkan-populations-based-autosomal-haploid-markers

Wikipedia 8 27 Nov 06:56 UTC

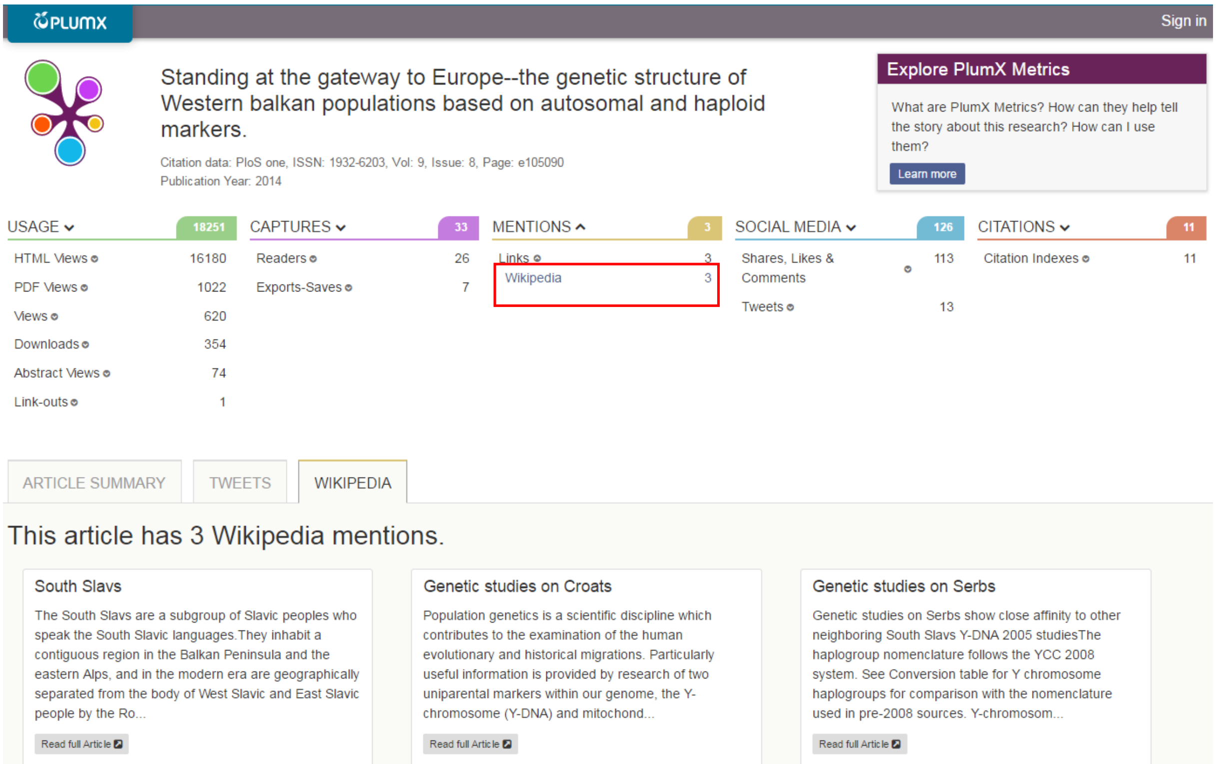
Wikipedia | Further Information

{"title"=>"Genetic studies on Serbs", "url"=>"http://en.wikipedia.org/wiki/Genetic_studies_on_Serbs", "timestamp"=>"2017-10-18T17:59:45Z"}
{"title"=>"Genetic studies on Croats", "url"=>"http://en.wikipedia.org/wiki/Genetic_studies_on_Croats", "timestamp"=>"2017-10-25T16:15:20Z"}
{"title"=>"South Slavs", "url"=>"http://en.wikipedia.org/wiki/South_Slavs", "timestamp"=>"2017-11-24T13:58:35Z"}
{"title"=>"File:Plos Balkans.png", "url"=>"http://commons.wikimedia.org/wiki/File:Plos_Balkans.png", "timestamp"=>"2017-10-02T05:57:41Z"}
{"title"=>"File:Region-wise FST-distances based on the variation of autosomal SNPs .png", "url"=>"http://commons.wikimedia.org/wiki/File:Region-wise_FST-distances_based_on_the_variation_of_autosomal_SNPs_.png", "timestamp"=>"2016-10-23T14:18:50Z"}
{"title"=>"File:PCA of the variation of autosomal SNPs in Western Balkan populations in Eurasian context.png", "url"=>"http://commons.wikimedia.org/wiki/File:PCA_of_the_variation_of_autosomal_SNPs_in_Western_Balkan_populations_in_Eurasian_context.png", "timestamp"=>"2016-10-23T14:23:24Z"}
{"title"=>"File:Plos.Balkans.2.png", "url"=>"http://commons.wikimedia.org/wiki/File:Plos.Balkans.2.png", "timestamp"=>"2017-10-02T02:50:10Z"}
{"title"=>"File:FST-distances based on the variation of autosomal SNPs of Western Balkans populations in a global context.png", "url"=>"http://commons.wikimedia.org/wiki/File:FST-distances_based_on_the_variation_of_autosomal_SNPs_of_Western_Balkans_populations_in_a_global_context.png", "timestamp"=>"2016-10-30T00:52:25Z"}

http://alm.plos.org/works/doi.org/10.1371/journal.pone.0105090



https://www.altmetric.com/details/2622786/wikipedia



https://plu.mx/plum/a/?doi=10.1371%2Fjournal.pone.0105090&display-tab=artifact-wikipedia#LINK_COUNT

Figure 7. Examples of different Wikipedia mentions across different altmetric aggregators: Lagotto, altmetric.com, and Plum Analytics are presented orderly (accessed on 29 November 2017).

3.3.3. Form of updates

It is not possible to know how exactly each aggregator queries the original social media sources and how often they update their data, besides the information reported by them. However, it is technically possible that differences in the date and time when social media events occurred and when the aggregator collected them, together with time lags in the frequency of updates of each aggregator, also cause discrepancies in the metrics provided by each aggregator. Although it is assumed that in most cases all aggregators have updated their platforms in real time or, depending on the data source, on a daily basis (as presented in Table 8), in most cases the information on the exact time of update across different aggregators is not available. Another reason for discrepancies in the updates of Mendeley metrics includes the time lags between the actual act of saving a paper by a Mendeley user, the update of the Mendeley readership of the publication by Mendeley and the moment when the aggregators collect their data. Moreover, the periodical update of the Mendeley database which only happens instantly for readership counts but periodically for their decrease[39]could also contribute to explain some of the differences found (see Appendix 1, Text 4).

[39] According to William Gunn (Director of Scholarly Communications in Mendeley), "When users delete their account and all their documents, the readership of that document doesn't change, until the batch clustering process is re-run and the new number of metadata records is generated. The same applies when a user deletes

3.3.4. Other technical reasons

Other technical issues include the matching rate of identifiers with journal publisher's platforms and their policy in allowing access, API speed, and rate of querying. Metrics depend on the matching rate of DOIs and URLs of an object by aggregators. There are differences across journal publisher platforms in resolving DOIs to journal landing pages. Whether a publisher allows DOIs resolving and how simple is this process (cookies problems, access denies, redirects) depend on the publishers' policies (Fenner and Lin, 2013). Hence, differences in metrics including any possible agreements between altmetric aggregators and specific publishers result in their different coverage of publications from different publishers (Carpenter, 2017). For instance, whether all the variations of journal publisher's URLs for a given DOI is known by the aggregator or not and the extent to which an aggregator is able to call the provider's API (for example Facebook API) for a given DOI to cover all the mentions across multiple URLs could influence aggregator's coverage of different publishers. Plum Analytics and Lagotto both track all the possible URLs for a given DOI from different publishers in both public and private posts. Thus, this can also contribute to the highest Facebook counts reported by Plum and Lagotto. Other issues such as availability of different ranges of identifiers (DOIs, PubMed, SSRN, ArXiv IDs, etc.) tracked, how shortened URLs are handled, how rate limits of data aggregator and third party provider APIs are handled, or the functioning of the rate of traffic over the API, are all technical issues that could influence the rate of querying APIs[40] and hence could also influence the metrics provided by the aggregators.

3.4. Conclusions

The proliferation of new social-media-based indicators has opened the possibility to study the interactions between social media and science in what can be seen as the social media studies of science (Wouters, Zahedi & Costas, 2018; Costas, 2017). However, the development of these studies has a strong dependency on the specific data and metrics available. Several grand challenges have been already pointed out regarding the development and potential applicability of these new data sources. "Heterogeneity", "data quality", and external "dependencies" have been argued as major challenges of altmetric data (Haustein, 2016). In this study, we specifically focus on the challenge related with "data quality" (although to some extent we also exemplify some of the "external dependencies" involved). Social media metrics data collection relies on a large range of different methodological and technical choices (e.g., APIs, identifiers tracked, forms of querying original sources, types of events recorded, selections of publishers) and reporting choices (e.g., aggregation of different types of counts into one single metric, grouping of different metrics into broader categories, combination of

a record from their library. In summary, the count of records can increase nearly instantaneously, but only decreases periodically" see:
www.niso.org/apps/group_public/view_comment.php?comment_id=632
www.niso.org/apps/group_public/view_comment.php?comment_id=610

[40] www.eventdata.crossref.org/guide/sources-in-depth/

different counts for different identifiers). Hence, it is important to understand how these choices may affect the data collected and reported by different aggregators.

This study describes how the social media metrics collected for a same set of DOIs at the same time may vary across different major altmetric aggregators. Similar results have been found in recent studies comparing different altmetric data aggregators (Meschede & Siebenlist, 2018; Ortega, 2017; Bar-Ilan and Halevi, 2017) as well as other previous studies (Chamberlain, 2013; Zahedi, Fenner, and Costas, 2014; 2015; Jobmann et al., 2014). For instance, the same high consistency across aggregators regarding Mendeley readership has been highlighted in these previous studies.

More specifically, our results showed that Lagotto and Plum Analytics provide the highest values of Mendeley readership. This can be explained by the choice of aggregating the counts coming from different identifiers of the same paper, or the different consideration of forms of readership (e.g., individual readership and group readership). Altmetric.com provides the highest value of tweets, which could be explained by the tracking and combination of counts from different versions of the same object. Plum Analytics provides the highest value of Facebook counts as it combines different events from Facebook in the same score, and CrossRef ED provides the highest value of Wikipedia mentions, as it collects mentions from different languages and edits of the same Wikipedia entry. Correlation analysis showed that the differences across aggregators for Mendeley readership counts are the least problematic, since the different values tend to correlate quite strongly (although the limited coverage of Altmetric.com with respect to Mendeley readership needs to be reminded. Although some relatively moderate correlations found across some data aggregators (particularly for CrossRef ED data with the other aggregators), the overall correlation analyses of Twitter counts suggest a reasonably good agreement among data aggregators. The lowest correlations among aggregators are found for Facebook and Wikipedia counts. For these sources it seems that the choices adopted by each of the aggregators in collecting and processing the counts have a strong relevance on the final counts reported by them. For these two sources, it is important for the users to understand what the aggregators are actually computing.

Overall we can argue that most of the differences found across data aggregators are explained by specific choices on the data collection and aggregation approaches as explained here. All of these choices can have different effects on the results and analytical approaches based on altmetric data. For example, the choice of aggregating all Mendeley readership from the different version of the same paper may have an inflationary effect. This inflationary effect can be challenging when the pairing of document identifiers is wrong (e.g., users wrongly linking DOIs and PMIDs [see Figures 1 and 2]). The choice of counting together different acts from the same social media source, like tweets or retweets, has also conceptual repercussions, since a tweet can be seen as an act of greater engagement (Haustein, Bowman & Costas, 2016) than a retweet. This can also be argued for the combined count of Facebook posts, shares, likes, etc., which breaks the internal homogeneity of the indicator (Wouters, Zahedi & Costas, 2018). This hinders the interpretability and meaning of the indicator, and opens the possibility for its easier manipulation. In a similar fashion, the counting of Wikipedia mentions of

different edits of the same Wikipedia entry has conceptual issues. The consideration of some different language versions of the same Wikipedia entry may be tricky, creating biases favoring publications form the countries of these languages (e.g., Finland and Sweden in the case of Altmetric.com Wikipedia counts and Spanish and Portuguese in the case of Plum Analytics Wikipedia counts). Also, the mere translations of a Wikipedia entry may derive in multiple mentions for a publication without reflecting a real engagement from the translators with the cited publications. Wikipedia articles are often translated by bots or applications.[41] and hence it doesn't always reflect the engagement of the translator with the content.
All in all, it is difficult to claim that some choices are better than others and hence the solution to the discrepancies among different aggregators cannot just be solved by recommending to use those that provide the higher counts. Actually, our results suggest that it is difficult to come up with universal recommendations on what aggregators must choose. All of them exhibit advantages and disadvantages depending on the choices and sources tracked. At best, we can talk about *overall recommendations* for both data aggregators and social media metrics users:

1. *Increase the transparency around the methodological choices in data collection*. Aggregators use different strategies in collecting, calculating, or updating metrics. These strategies may involve technical issues such as linking duplicate records or merging different acts from the same platform in one count. In this sense, although the efforts of data aggregators for making better aggregations of metrics for publications is commendable, it is critical that they are more transparent on how they have collected and aggregated the data. Users, should also be aware that these choices may imply potential risks for their analysis, and should demand more and transparent explanations on these for the analytical use of the data.

2. *Increase the awareness of unintended effects of methodological choices*. Both users, researchers, and data aggregators should be aware of the unintended effects that methodological choices can have in the use and application of social media metrics data and its application. For instance, tracking mentions to the publisher's URL (in addition to some other identifiers) may increase the counts to publications from some publishers (which are tracked) but miss those of other less known publishers, thus creating a bias towards the tracked publishers. Hence, it could be argued that Crossref ED approach of just tracking registered DOIs could be seen as a less biased approach, although it provides lower Twitter scores than other aggregators. Similarly, if some metrics or counts are dependent on the occurrence of other events (e.g., Mendeley counts in Altmetric.com) this can also have effects on the analytical validity of this data source depending on the objectives of the user. These potential biases and limitations should be explained and users must be aware of them.

[41] https://en.wikipedia.org/wiki/Swedish_Wikipedia

3. *Increase the transparency around the computation of different social media acts*. Disclosing the combined computation of different social media acts into one metric is also important (e.g., whether Facebook or Twitter counts include posts, likes, shares, comments). This is critical in order to understand the internal homogeneity and conceptual value of the metrics reported. Considering the infancy of social media metrics and the uncertainty in the relationship and meaning of the different social media events, it seems reasonable to argue that keeping events separate as much as possible is probably the best approach from an analytical perspective. The combination of conceptually different metrics into one single measure may introduce misunderstandings, misuses, and even manipulations that could have negative effects on the further application of social media metrics.

4. *Increase the replicability and interactivity of the data reported.* Recording, disclosing, and making available the original raw social media data to the users would allow them to make their own choices, as well as to obtain a better idea of the origin and provenance of the data collected. As emphasized by the *NISO altmetrics data quality Code of Conduct*, documenting the degree of transparency on how each aggregator queries different data sources and the processes taken, is to be preferred in order to make it possible to verify the metrics. Also, providing some information on the accessible content by the different data sources, characteristics of their underlying data, internal processes applied by aggregators, how they access external sources, and their strategy to calculate or report metrics, helps to gain better understanding of the relevant issues faced by each aggregator when collecting and processing social media data. From the user point of view, it is recommended to demand more interactive possibilities when it comes to the use and analysis of social media metrics. Hence, incorporating analytical features through which users can choose sources, periods of time, types of social media acts as well as indicators, can help to empower the user in the application, replication, and better interpretation of social media metrics.

The results of this paper provide some original insights about current data challenges in social media metrics. It is important to emphasize that the validity and reliability of social media metrics sources should be constantly checked and discussed, particularly among altmetric data aggregators, researchers in social media metrics, and the users of this data. The importance of these methodological choices in data collection and calculation of metrics should be incorporated in the overall discussion around social media metrics research. Understanding how methodological and technical choices can influence the analytical reliability and validity of social media metrics is a critical element in the future development of social media studies of science. Future research should also focus on providing further insights and possible solutions for current and potential data challenges in social media data collection. Also, other suggestions for future studies could include whether the choice of journals (open access vs. closed) affects the results obtained in the current study. Moreover, the extent to which each aggregator combines metrics for the different versions of publications, different identifiers of the same object, and how does these combinations influence the metrics provided by the aggregators needs to be further studied.

3.5. Acknowledgement

The idea of this study received the first altmetrics project funding award at the 1:AM altmetrics conference, September 2014, London (UK). The authors are grateful for the fundings awarded by Thomson Reuters/Altmetric.com. We also acknowledge funding from the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP). Special thanks to Euan Adie from Altmetric.com, and Alley Pollen and Stefanie Faulkner from Plum Analytics for their useful feedback and remarks on this paper.

3.6. References

Bar-Ilan, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271. https://doi.org/10.1007/s11192-008-0216-y

Bar-Ilan, J. & Halevi, G. (2017). Altmetric Counts from Different Sources: A Case Study of Journal of the Association for Information Science and Technology (JASIST) Articles Published Between 2001 and Mid 2017. *The 2017 Altmetrics Workshop: The Dependencies of Altmetrics*, 26 September 2017, Toronto, Canada.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY*, 30(1), 89–116. https://doi.org/10.1007/s00146-014-0549-4

Carpenter, T. (2017, February 2). Plum Goes Orange: Elsevier Acquires Plum Analytics. [blog post]. The Scholarly Kitchen. Retrieved from: https://scholarlykitchen.sspnet.org/2017/02/02/plum-goes-orange-elsevier-acquires-plum-analytics/

Chamberlain, S. (2013). Consuming article-level metrics: Observations and lessons from comparing aggregator providers data. *Information Standards Quarterly*, 25 (2), 4-13.

Costas, R. (2017). Towards the social media studies of science: social media metrics, present and future. *Biblioteca*s Anales de Investigación, 13(1), 1–5.

Delgado López-Cózar, E., Robinson-García, N., & Torres-Salinas, D. (2014). The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, 65(3), 446–454. https://doi.org/10.1002/asi.23056

Fenner, M. (2013, 13 October). Challenges in automated DOI resolution [Blog Post]. Retrieved from: http://blog.martinfenner.org/2013/10/13/broken-dois/

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

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation? Review of the Literature. *Journal of Informetrics*, 11(3), 823–834 https://doi.org/10.1016/j.joi.2017.06.005

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. https://doi.org/10.1007/s11192-015-1798-9

Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics,* 108(1), 413-423.

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting Altmetrics: Viewing Acts on Social Media through the Lens of Citation and Social Theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication* (pp. 372–406). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110308464-022

Jacsó, P. (2011). Google Scholar duped and deduped: the aura of “robometrics.” *Online Information Review*, 35(1), 154–160 https://doi.org/10.1108/14684521111113632

Jobmann, A., Hoffmann, C. P., Künne, S., Peters, I., Schmitz, J., & Wollnik-korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie – Praxis Und Forschung*, 3(1), 1–19.

Lin, J., Fenner, M. (2013). Altmetrics in evolution: Defining and redefining the ontology of article-level metrics. *Information Standards Quarterly*, 25 (2) (2013), pp. 20-26 https://doi.org/10.3789/isqv25no2.2013.04

Meho, L. I., & Sugimoto, C. R. (2009). Assessing the scholarly impact of information studies: A tale of two citation databases: Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 60(12), 2499–2508. https://doi.org/10.1002/asi.21165

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. *Scientometrics*. https://doi.org/10.1007/s11192-018-2674-1

Moed, H. F., Bar-Ilan, J., & Halevi, G. (2016). A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics*, 10(2), 533–551. https://doi.org/10.1016/j.joi.2016.04.017

National Information Standards Organization. (2016). Outputs of the NISO Alternative Assessment Metrics Project. Retrieved from: http://www.niso.org/apps/group_public/download.php/17091/NISO RP-25-2016 Outputs of the NISO Alternative Assessment Project.pdf

Olensky, M., Schmidt, M., & van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and IFQ in comparison to the Web of science. *Journal of the Association for Information Science and Technology*, 67(10), 2550–2564. https://doi.org/10.1002/asi.23590

Ortega, J. L. (2017). Reliability and accuracy of altmetric providers : a comparison among Altmetric, PlumX and Crossref Event Data. Retrieved from: https://osf.io/fuekb/

Peters, I., Kraker, P., Lex E, Gumpenberger, C., and Gorraiz, J.I. (2017). Zenodo in the Spotlight of Traditional and New Metrics. *Frontiers Research Metrics*. Anal. 2:13. https://doi.org/10.3389/frma.2017.00013

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., Costas, R. (2014). New Data, New Possibilities: Exploring the Insides of Altmetric.Com. *El Profesional de la Information*, 23(4), 359–366. https://doi.org/10.3145/epi.2014.jul.03

Valderrama-Zurián, J. C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics*, 9(3), 570–576. https://doi.org/10.1016/j.joi.2015.05.002

Wass, J. (2017, 16 August). You do want to see how it's made – seeing what goes into altmetrics. [Blog Post]. Retrieved from: http://altmetricsconference.com/you-do-want-to-see-how-its-made-seeing-what-goes-into-altmetrics/

Wouters, P., & Costas, R. (2012). *Users, narcissism and control – tracking the impact of scholarly publications in the 21st century*. Utrecht: SURF foundation. Retrieved from http://www.surf.nl/nl/publicaties/Documents/Users narcissism and control.pdf

Wouters, P., Zahedi, Z., & Costas, R. (2018). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of Quantitative Science and Technology Research*. Springer.

Zahedi, Z., Bowman, T. D., & Haustein, S. (2014a). Exploring data quality and retrieval strategies for Mendeley reader counts. In *SIG/MET workshop*, ASIS&T 2014 annual meeting, Seattle, US.

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs. In *altmetrics 14 Workshop and the Web Science Conference*, Bloomington, USA. https://doi.org/ 0.6084/m9.figshare.1041821

Zahedi, Z., Fenner, M., & Costas, R. (2015). Consistency among altmetrics data providers/aggregators: What are the challenges? *In altmetrics15: 5 years in, what do we know?* Amsterdam, The Netherlands.

CHAPTER 4

On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications[1]

[1] This chapter is based on:
Zahedi, Z., and Haustein, S. (2018). On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications. *Journal of Informetrics*, 12 (1): 191-202. http://doi.org/ 10.1016/j.joi.2017.12.005

Abstract

In this chapter we present a first large-scale analysis of the relationship between Mendeley readership and citation counts with particular documents' bibliographic characteristics. A data set of 1.3 million publications from different fields published in journals covered by the Web of Science (WoS) has been analyzed. This work reveals that document types that are often excluded from citation analysis due to their lower citation values, like editorial materials, letters, or news items, are strongly covered and saved in Mendeley, suggesting that Mendeley readership can reliably inform the analysis of these document types. Findings show that collaborative papers are frequently saved in Mendeley, which is similar to what is observed for citations. The relationship between readership and the length of titles and number of pages, however, is weaker than for the same relationship observed for citations. The analysis of different disciplines also points to different patterns in the relationship between several document characteristics, readership, and citation counts. Overall, results highlight that although disciplinary differences exist, readership counts are related to similar bibliographic characteristics as those related to citation counts, reinforcing the idea that Mendeley readership and citations capture a similar concept of impact, although they cannot be considered as equivalent indicators.

Keywords

Mendeley readership; WoS Citation; Bibliographic characteristics; Document types

4.1. Introduction

4.1.1. Effect of document characteristics on citation impact

Measuring research impact using citation analysis has a long tradition in the field of scientometrics. Today, citation-based indicators are widely used and play a central role in the evaluation of scientific works. Despite their *de facto* use as proxies of scientific quality, citations are not able to fully capture the use and influence of scientific papers (Moed, 2005; MacRoberts & MacRoberts, 2017). Bibliometric research has also shown that a variety of factors can influence citation counts (Opthof & Leydesdorff, 2010; Waltman et al., 2011; Larivière & Gingras, 2011). Such factors include, the document types and age of publications, their number of pages, the length of their titles and reference lists (Bornmann & Leydesdorff, 2015; Bornmann, Leydesdorff, & Wang, 2014; Vieira & Gomes, 2010); their different theoretical or methodological approaches (Antonakis et al., 2014); whether they are open access (Hajjem, Harnad, & Gingras, 2006); the citation propensity of their fields and their interdisciplinarity (Yegros-Yegros, Rafols, & D'Este, 2015); or the Impact Factor of their publication journal (Boyack & Klavans, 2005).

Numerous previous studies have analyzed whether citation impact is affected by various document characteristics. These studies have explored different characteristics at the article, journal, and author levels using correlation and regression analyses. For example, in the Natural, Life, and Health sciences (Thelwall, 2017), papers with unusual and obscure titles were associated with lower citation impact. Mixed results were found regarding the effect of title length (Stremersch et al., 2015; Jacques & Sebire, 2010), or titles that included non-alphanumeric characters such as hyphens or colons (Buter & Van Raan, 2011; Haslam et al., 2008; Nair & Gibbert, 2016). Based on the assumption that longer articles with longer reference lists may reflect in-depth analysis and diversity of ideas, the number of pages and references have also been analyzed as factors that may affect citation counts (Fox & Boris, 2016). The results showed that papers with more references and more pages tended to get more citations (Ajiferuke & Famoye, 2015; Davis et al., 2001). Similarly, the number of authors, institutes, and countries involved in a given publication may indicate the extent of collaboration, which is again assumed to increase citation impact. However, results regarding the effect of collaboration on citation rates are mixed (for an overview see Onodera & Yoshikane, 2015) as regards variations by country of collaboration (Thelwall & Sud, 2016), level of collaboration (e.g., whether national, international, intra/inter institutional) (Leimu & Koricheva, 2005), or authors and disciplines (Williams et al., 2009). For a recent review of studies analyzing factors affecting citation counts we refer to Tahamtan, Safipour Afshar, & Ahamdzadeh (2016).

4.1.2. Effect of document characteristics on social media visibility

In the context of recently introduced altmetrics—or, more specifically, its subset of social media based metrics (e.g., Facebook, Twitter, blogs, Wikipedia, Mendeley)—the effect to which some factors influence social media activity remains understudied. One large-scale study examining the effect on social media metrics of typical document characteristics (including document type, discipline, number of pages, title length, number of references, and collaboration patterns) conducted by Haustein, Costas, & Larivière (2015). This study was based on Altmetric.com and Web of Science data and found that although effects were weaker than for citations, documents were more likely to be tweeted if they had longer reference lists and involving a greater number of authors, institutes, and countries. Correlations between social media metrics and document characteristics were, however, quite low to non-existent, which was mostly due to the skewed nature of social media events related to journal articles, with most of them having no metrics at all. Social media metrics (particularly Facebook and Twitter counts) correlated mostly among each other, indicating a circular relationship (Bourdieu, 1998), meaning that being picked up by one social media increases the chances of being picked up by another one. Haustein et al. (2015) also found that news items and editorials were among the most tweeted document types, which indicates that outputs that contain more condensed, novel, opinion-based and easy-to-understand pieces tend to be more popular on Twitter. The results contrast with the citation patterns for these types of documents, which are substantially less cited than articles and reviews. Overall, the study by Haustein et al. (2015) showed that characteristics that typically are related to higher citation counts had a smaller relationship with social media counts, sometimes even in an entirely different manner (for instance, longer titles were associated with higher citation counts but with lower Twitter mentions).

4.1.3. Mendeley readership and citation counts

Mendeley is an online reference manager that allows users to save documents in their own libraries and share their libraries with others. Statistics about how often a particular document is saved are made available via the Mendeley API as 'readership' counts. While this count is described by Mendeley as 'readership', it does not actually indicate that the user who saved the document has actually 'read' it, but simply that the user has saved the reference in the library. As such, Mendeley 'saves' are seen more as acts of access to documents than of their appraisal (Haustein, Bowman & Costas, 2016), indicating that the level of engagement captured by these acts is very low.

However, Mendeley has been identified as the most prevalent and noteworthy altmetric source. It has been found that readership counts often exceed citations, and that there is a high representation of recent publications on the platform (Thelwall & Sud, 2015). Compared to other altmetric indicators, Mendeley readership counts were shown to have moderate to strong correlations with citation counts (for a review see Sugimoto et al., 2016), which reflects a greater similarity with citations than other altmetric indicators (Costas, Zahedi, & Wouters, 2015a). This can be explained by the large numbers of academic users in Mendeley, and the frequent use of Mendeley in a pre-citation context (Mohammadi, Thelwall, & Kousha, 2016). The number of Mendeley users who have added an article to their libraries has been suggested as an early indicator of citation impact (Thelwall & Sud, 2015), and Mendeley itself has been identified as a relevant tool with which to identify highly cited publications (Zahedi, Costas,

&Wouters, 2017). Mendeley readership distributions have also been shown to be very similar to citation distributions (Costas, Haustein, Zahedi, & Larivière, 2016), and it has been suggested that field-normalized readership scores could be calculated in a similar fashion as for citations (Bornmann & Haunschild, 2016).

Given these similarities between readership and citations, one might expect that Mendeley readership counts are also related to the same document characteristics as citations. While some characteristics, such as document age (Thelwall, Haustein, Larivière, & Sugimoto, 2013), disciplines (Haustein et al., 2015), topics (Costas, Zahedi, & Wouters, 2015a), or countries (Alperin, 2015) have been already explored for Mendeley, a systematic study of other quantitative document characteristics previously investigated for citations is still lacking in the literature regarding Mendeley readership. A recent study (Didegah, Bowman, & Holmberg, in press[1]) investigated the relationship between some factors (such as JIF, cited references, title length, country and institute's prestige, and field type and size) with Mendeley and citation counts for a sample of Finnish WoS papers. However, a global large-scale disciplinary analysis of the relationship of these characteristics as well as some other factors is still missing in the literature.

The present work represents the first large-scale analysis of the relationship between Mendeley readership and specific documents' bibliographic characteristics. Specifically, this study aims to improve the understanding of the relationship between Mendeley readership and selected document characteristics, including document types, number of pages, title length, length of reference list, and number of authors, institutes, and countries of the papers. We study how the relationship between these bibliographic characteristics and readership is similar to and/or differs from that observed for citations and, how this relation varies across different fields. The selection of document characteristics represents only a limited number of quantitative variables, and we acknowledge that it does not consider other qualitative aspects that might affect the extent to which articles attract users on Mendeley. Nevertheless, this study will contribute to a better understanding of how Mendeley readership relates to basic document characteristics by providing a clear framework of this relation. This could contribute to the identification of document-related differences between Mendeley readership and citations that can help the future construction of appropriate and meaningful indicators based on Mendeley readership. The study builds on the work by Haustein et al. (2015) and Zahedi et al. (2016) by studying the same document characteristics, while taking into account longer citation and readership windows. Also, this study improves on these previous studies by including Mendeley readership counts and documents from several different disciplines and by using more advanced regression analysis (in contrast with the more basic correlation analysis employed in previous studies). The paper addresses the following research questions:

[1] This study uploaded to arXiv (https://arxiv.org/abs/1710.08594) while our paper was under review by the Journal of Informetrics.

- To what extent do document type and document characteristics (i.e., number of pages, length of title and reference list, and number of authors, institutes and countries involved) associate with the number of Mendeley readership?
- How do these relationships between document characteristics and Mendeley readership vary across disciplines? And how do they compare with those observed for citations?

4.2. Data and Methodology

This study compares Mendeley readership to citation counts received by 2012 Web of Science (WoS) publications with a Digital Object Identifier (DOI) from all disciplines (n=1,339,279). Citation counts from the Centre for Science and Technology Studies (CWTS) in-house WoS database were collected through the end of August 2015. Mendeley readership counts were extracted from the Java Script Object Notation (JSON) files obtained from querying the Mendeley Application Programming Interface (API) using DOIs in July 2015. The analyzed document properties included document type (as recorded by WoS), number of pages, number of cited sources in the reference list (including non-source items), number of characters in the title, number of authors, institutes, and countries of the paper, as well as the scientific disciplines (according to the Leiden Ranking (LR) classification based on the CWTS in-house version of WoS).

General descriptive statistics were computed for all documents (n= 1,339,279). This included the percentage of papers with at least one citation or one Mendeley readership count (coverage) and the average number of counts per paper (density). To assess the influence of each of the independent variables (i.e., title length, number of pages, number of references, authors, institutes and countries) on readership and citation counts, a standard linear regression (Ordinary Least Squares) analysis was performed using RStudio for all articles and reviews (n= 1,197,162). The OLS analyses were performed to detect if there were any disciplinary differences in the factors influencing readership vs. citation impact[2]. Although there are some debates surrounding how to choose a proper regression model for skewed distributions (see Ajiferuke & Famoye, 2015), the OLS analysis (after log-transforming readership and citation counts[3]) has been shown to be the most suitable regression strategy for citation data and altmetrics (Thelwall & Wilson, 2014). This method proved to be more reliable than negative binomial (NB) or zero inflated negative binomial (ZINB) regression analyses (Thelwall & Wilson, 2014) as it takes into account very high values which are typical for skewed distributions (Thelwall, 2016), which is similar to the skewed log normal distributions of both citation and readership counts in this study.

[2] Only total readership counts (not the readership disaggregated by Mendeley users) are included in the regression model.

[3] The value of 1 was added to readership and citation counts to include papers without zero values in the regression.

4.3. Results

4.3.1. Distribution of citation and readership across different disciplines

Out of the 1,339,279 WoS publications, 93.6% (n= 1,254,852) were found in Mendeley via the DOI (Table 1). Of the total publications, 81.7% (n= 1,094,166) were cited at least once in WoS and 84.2% (n= 1,127,849) had at least one readership (at time of data collection). This set of publications accumulated 14,732,103 readership counts and 10,289,891 citation counts in total. Table 1 presents the most important results regarding the coverage and density of both counts in our database. On average, each paper received approximately 7.7 citations and 11 readers, which indicates that documents have been saved more in Mendeley than cited. Regarding the share of publications with at least a Mendeley readership and a citation, they are very close in most disciplines with the exception of the Social sciences and humanities, in which 81.7% of publications had been saved on Mendeley, while only 64.1% had been cited (Table 1).
On average, papers from the Social sciences and humanities had 14.1 readership counts on Mendeley and were cited only 4 times in WoS. The Social sciences and humanities represent the discipline with the second highest readership density after Life and earth sciences (16.5 mean readership vs. 8.6 mean citation counts) and Biomedical and health sciences (12.2 mean readership vs. 8.7 mean citation counts). In Natural sciences and engineering (8.8 mean readership vs. 9 mean citation counts) and Mathematics and computer science (7.8 mean readership vs. 4.4 mean citation counts) Mendeley readership counts are lower. Natural sciences and engineering is the only discipline where average citation counts (9) slightly exceed average Mendeley readership counts (8.8) (Table 1).

Table 1. Mendeley and citation coverage and density (standard deviation) per discipline.

		All disciplines	**Biomedical & health sciences**	**Life & earth sciences**	**Mathematics & computer science**	**Natural sciences & engineering**	**Social sciences & humanities**
		N=1,339,279	N=595,254	N=254,817	N=135,445	N=413,862	N=159,389
Citations	Coverage	81.7	83.4	89.1	75.4	87.3	64.1
	Density	7.7	8.7	8.6	4.4	9.0	4.0
	Std. dev.	*18.8*	*20.9*	*19.0*	*12.4*	*21.3*	*8.7*
Mendeley Readership	Coverage	84.2	86.5	91.4	76.4	83.7	81.7
	Density	11.0	12.2	16.5	7.8	8.8	14.1
	Std. dev.	*23.9*	*26.5*	*32.5*	*29.5*	*18.0*	*27.0*

4.3.2. Analysis of relation between document types and their citation and readership impact

Figure 1 presents the coverage and average values of both citation and readership for the most important document types in WoS. Reviews and articles are the document types that were most commonly cited and saved on Mendeley. This is to be expected since these types of publications represent the most important research findings that are widely relevant to various audiences. Although both coverage and density were higher for reviews and articles, editorial materials (6.5 density; 68.2% coverage), and news items (5 density; 59.4% coverage) were also frequently saved by Mendeley users, followed by letters (2.7 density; 58.4% coverage), book reviews (1.5 density; 28.8% coverage), biographical items (0.8 density; 28.4% coverage), corrections (3.4 density; 25.4% coverage), and meeting abstracts (0.6 density; 24.7% coverage).

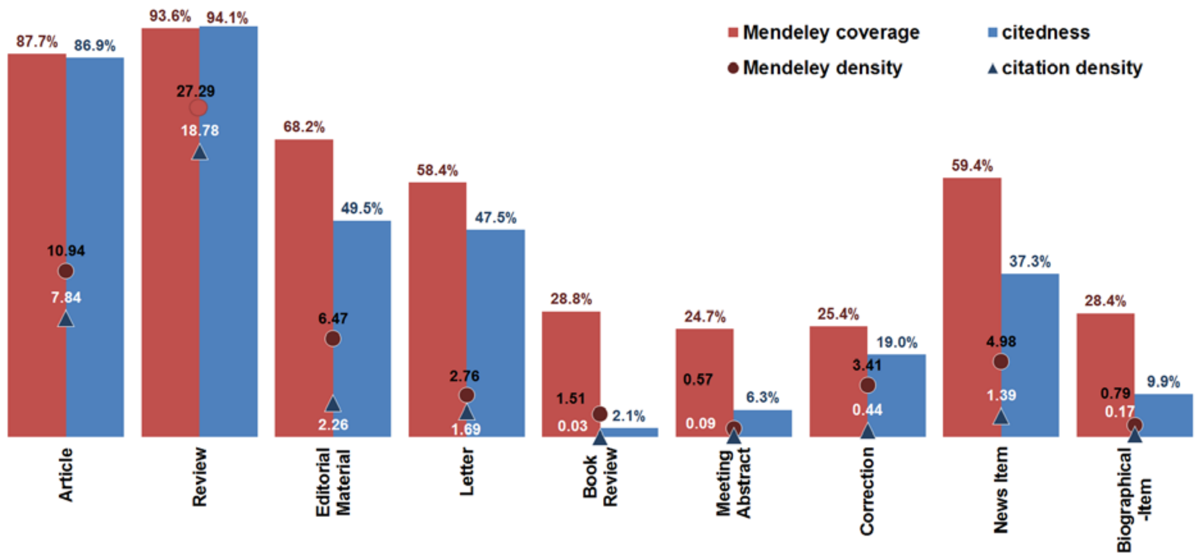


Figure1. Mendeley coverage and density, citedness and citation density per document type. Document types appear in order of frequency: article (N=1,132,428), review (N=64,734), editorial material (N=60,533), letter (N=29,410), book review (N=21,710), meeting abstract (N=13,071), correction (N=9,817), news item (N=4,880) and biographical item (N=2,302). Document types that occurred less than 2,300 times were not analyzed.

4.3.3. Analysis of relations between document characteristics and citation and readership impact across disciplines

General average values and standard deviation of the examined variables are presented in Table A1 in the Appendix 2. This table illustrates that, on average, papers from the Life and earth sciences, Biomedical and health sciences, and Natural sciences and engineering tend to have the longest titles. Also, these papers are more collaborative than those from both Social sciences and humanities and Mathematics and computer science in terms of number of authors involved. Papers from Social sciences and humanities have the longest reference lists and number of pages.

The results of the OLS regression analysis among the dependent and independent variables of articles and reviews (n=1,197,162) and their individual disciplines[4] are presented in Table 2. Multicollinearity was tested before making inferences about the coefficient estimates using

[4] Bootstrap and 95% confidence interval showed significant results although the power (R-squared) is not very strong which is common in social sciences studies. This means that although variables in the model have statically significant coefficients, they account for only a small portion of the variation in the each dependent variables separately.

Variance Inflation Factor (VIF). VIF is an indicator of how much the variance of estimated coefficients change due to multicollinearity, although there is no consensus of what is a high VIF value. VIF value above 10 is often regarded as an indicator of serious multicollinearity (Tabachnick & Fidell, 2007) and a VIF value above 4 is typically considered problematic (Rovai, Baker, & Ponton, 2013). As a result, the number of authors and institutes from Natural sciences and engineering and their interpretations have been excluded from the OLS model due to their high VIF values (>10). Number of institutes in all other fields have VIF values between 2.1 to 2.7[5], while all other variables have VIF values below 1.9; hence no collinearity is expected.

Table 2. Ordinary Least Squares regression analysis (OLS) between dependent (citation and readership) and independent (title length, number of references, pages, authors, institutes, and countries) variables across all articles and reviews from their individual LR fields.

a) Dependent variable Log (Citation +1)

Biomedical & health sciences (n= 502,610)

	Coefficients (B)	B	Std. Error	CI 95% (B)		VIF
Number of references	0.010***	0.311	0.000	0.009	0.010	1.309
Title length	0.001***	0.046	0.000	0.001	0.001	1.036
Number of pages	0.004***	0.027	0.000	0.004	0.004	1.295
Number of authors	0.033***	0.169	0.000	0.033	0.034	1.907
Number of institutes	-0.009***	-0.022	0.001	-0.011	-0.007	2.607
Number of countries	0.111***	0.097	0.002	0.108	0.115	1.670

Life & earth sciences (n=242,117)

	Coefficients (B)	B	Std. Error	CI 95% (B)		VIF
Number of references	0.011***	0.347	0.000	0.011	0.011	1.453
Title length	0.000***	-0.014	0.000	0.000	0.000	1.008
Number of pages	-0.014***	-0.094	0.000	-0.015	-0.013	1.444
Number of authors	0.037***	0.163	0.001	0.036	0.038	1.678
Number of institutes	-0.001†	-0.002	0.002	-0.004	0.002	2.651
Number of countries	0.090***	0.078	0.003	0.084	0.096	1.864

Natural sciences & engineering (n=404,457)

	Coefficients (B)	B	Std. Error	CI 95% (B)		VIF
Number of references	0.014***	0.401	0.000	0.014	0.014	1.274
Title length	0.002***	0.051	0.000	0.001	0.002	1.005
Number of pages	-0.017***	-0.106	0.000	-0.017	-0.016	1.288
Number of countries	0.076***	0.091	0.001	0.073	0.078	1.015

[5] Although the spearman correlations among the number of institutes and authors and the number of authors and countries are not very high ranging from .4 to .6 across fields (see Table A2 in the appendix), the results for number of institutes should be interpreted with caution (due to its higher VIF value than other variables).[6] The results should be interpreted with caution due to its higher VIF value (VIF=2.6) than other variables.

Mathematics & computer science
(n= 131,220)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.015***	0.312	0.000	0.015	0.016	1.148
Title length	0.002***	0.062	0.000	0.002	0.002	1.040
Number of pages	-0.008***	-0.073	0.000	-0.009	-0.008	1.162
Number of authors	0.038***	0.115	0.001	0.036	0.040	1.464
Number of institutes	0.006*	0.008	0.003	0.000	0.012	2.195
Number of countries	0.096***	0.062	0.005	0.086	0.106	1.703

Social sciences & humanities
(n=125,795)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.010***	0.312	0.000	0.010	0.010	1.224
Title length	0.001***	0.047	0.000	0.001	0.002	1.072
Number of pages	-0.018***	-0.150	0.000	-0.018	-0.017	1.265
Number of authors	0.092***	0.232	0.001	0.089	0.094	2.136
Number of institutes	0.026***	0.037	0.003	0.020	0.032	2.723
Number of countries	0.064***	0.043	0.005	0.055	0.074	1.644

b) Dependent variable Log (Readership +1)

Biomedical & health sciences
(n= 502,610)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Model						
Number of references	0.010***	0.302	0.000	0.010	0.010	1.309
Title length	-0.001***	-0.023	0.000	-0.001	-0.001	1.036
Number of pages	0.004***	0.026	0.000	0.004	0.005	1.295
Number of authors	0.005***	0.022	0.000	0.004	0.005	1.907
Number of institutes	0.010***	0.001	0.002	0.008	0.011	2.607
Number of countries	0.131***	0.106	0.002	0.127	0.135	1.670

Life & earth sciences
(n=242,117)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.013***	0.367	0.000	0.013	0.013	1.453
Title length	-0.002***	-0.076	0.000	-0.003	-0.002	1.008
Number of pages	-0.019***	-0.114	0.000	-0.020	-0.019	1.444
Number of authors	0.012***	0.044	0.001	0.010	0.013	1.678
Number of institutes	0.025***	0.040	0.002	0.021	0.029	2.651
Number of countries	0.132***	0.098	0.003	0.125	0.139	1.864

Natural sciences & engineering (n=404,457)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.012***	0.329	0.000	0.012	0.012	1.274
Title length	-0.001***	-0.025	0.000	-0.001	-0.001	1.005
Number of pages	-0.015***	-0.088	0.000	-0.015	-0.014	1.288
Number of countries	0.035***	0.040	0.001	0.032	0.038	1.015

Mathematics & computer science (n= 131,220)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.026***	0.447	0.000	0.026	0.027	1.148
Title length	0.000†	0.001	0.000	0.000	0.000	1.040
Number of pages	-0.018***	-0.132	0.000	-0.019	-0.017	1.162
Number of authors	0.048***	0.118	0.001	0.045	0.050	1.464
Number of institutes	0.017***	0.017	0.004	0.010	0.024	2.195
Number of countries	0.057***	0.031	0.006	0.045	0.069	1.703

Social sciences & humanities (n=125,795)

	Coefficients (B)	β	Std. Error	CI 95% (B)		VIF
Number of references	0.013***	0.342	0.000	0.013	0.013	1.224
Title length	0.002***	0.049	0.000	0.002	0.002	1.072
Number of pages	-0.024***	-0.175	0.000	-0.025	-0.023	1.265
Number of authors	0.067***	0.144	0.002	0.063	0.070	2.136
Number of institutes	0.017***	0.020	0.004	0.010	0.024	2.723
Number of countries	0.128***	0.072	0.006	0.117	0.139	1.644

Statistical significance: *p < .05; **p < .01; ***p < .001, † shows not significant coefficient (p>0.05); R^2 = amount of variance explained by IVs, B = Unstandardized coefficient, β = Standardized coefficient (values for each variable are converted to the same scale so they can be compared), SE = Standard Error, CI = Confidence Interval, VIF= Variance Inflation Factor.

Based on Table 2, number of cited references, authors, institutes, and countries associate with increased citation and readership counts across all fields. There are two exceptions: first, the relation between number of institutes is not significant (sig=0.55, p>0.05) in Life and earth sciences; second, in Biomedical and health sciences, number of institutes is not associated with high citation impact since the effect is very low[6] (Table 2). Longer papers associate with increased citation and readership counts in Biomedical and health sciences, while they associate with decreased citation and readership counts in all other fields. In terms of title length, papers with longer titles are cited more than those with shorter titles. This is supported by both the OLS analysis and the correlation analysis (see Table A2 in the Appendix 2), since citations are positively related to the number of characters of papers' titles. Yet, it seems that this is not the case for Mendeley readership, since OLS and correlations tend to be negative.

[6] The results should be interpreted with caution due to its higher VIF value (VIF=2.6) than other variables.

The only exception is seen for Social sciences and humanities, where papers with longer titles are saved more than those with shorter tiles in Mendeley. The relationship between title length and readership impact in Mathematics and computer science is not significant (sig=0.65, p>0.05).

4.4. Discussion and conclusions

In this paper, we present the results of a first large-scale analysis of the relationship between document characteristics and Mendeley readership and citation counts for a large set of WoS publications. The aim of this analysis was to test how certain document characteristics are related to citation and readership counts, and how these patterns differ across fields.

The results showed that papers are overall saved in Mendeley more than cited. Activity on social bookmarking platforms such as Mendeley is expected to exceed citations, given the shorter time lag and broader type of use of saving (reading) over citing—not all that is read is cited—as well as the conceptually larger audience of readers as compared to citing authors. Also, the process of saving documents in Mendeley involves less time and engagement than reading, understanding, and citing them in a paper. The higher density of readership than citation was particularly evident in Social sciences and humanities. This reflects how the citation culture of this discipline is strongly influenced by the coverage of WoS, which excludes citations from books and regional and non-English language journals, that play central role in the Social sciences and humanities. Moreover, the citation delay is particularly high in the Social sciences and humanities, so that only a fraction of citations will have appeared during the citation window of 40 months captured in this study. In contrast, Mathematics and computer science is the field with the lowest Mendeley readership density. The same low presence of Mendeley readers in Mathematics compared to other fields has also been reported in the study by Thelwall (in press). This might suggest that other reference managers such as JabRef, EndNote, or BibSonomy could be more popular in this field or that scholars from this field do not use any reference managers at all. Studies have found that in general uptake of Mendeley is low among researchers (Bowman, 2015; Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014; Van Noorden, 2014). The study by Haustein et al. (2015) also showed lower social media mentions from Twitter, Facebook or blogs for papers from Mathematics and computer science than those from Social sciences and humanities, Biomedical and health sciences, and Life and earth sciences. This supports the idea that social, health, and environmental topics receive more attention in social media than highly technical topics (Costas, Zahedi, & Wouters, 2015b).

Our study shows that reviews and articles are the most prevalent document types in Mendeley. This is to be expected, since these document types represent the most important research findings that are widely relevant to various audiences. However, compared with citations, Mendeley readership exhibits higher coverage and relatively higher density values for other document types (e.g., editorial materials, letters, and news items). This indicates that although they do not contain material that is frequently cited, these document types are informative and do attract Mendeley users. This supports the idea that Mendeley readership could be useful for the analysis and evaluation of these document types, especially given that most of them are traditionally excluded from citation analysis (Waltman et al., 2011) due to their lower citation frequency. The popularity of these other document types has been reported for other altmetrics, particularly for Twitter (Haustein et al., 2015) and research blogs (Schema, Bar-Ilan, & Thelwall, 2015). This stronger presence of other document types in Mendeley also suggests that Mendeley readership may capture a mix of citation and social media patterns, since documents with different relevance for citations and social media are reasonably saved in Mendeley (e.g., news media items or editorial material).

Regarding the relationship between various document characteristics and Mendeley readership and citation counts, similar patterns were found in most cases. Akin to citations, papers with more references, written by many authors from several institutes and countries, are saved more frequently in Mendeley. This is because collaborative papers based on a large number of cited references may reflect more in-depth and diverse research of higher quality (Fox & Boris, 2016) or use of different expertise from different institutes and countries. However, the association of collaboration with citation counts varies across fields and years. For instance, number of (co)authors associates with a high citation impact in Arts and humanities, Chemistry, Pharmacology, Toxicology, Pharmaceutics (Thelwall & Sud, 2015), and Ecology (Fox & Boris, 2016). In some fields, such as Biomedicine, a positive association of the number of authors with citation impact exists until two or three years after publication and thereafter decreases (Bornmann & Leydesdorff, 2015). Country and level of collaboration (institutional, national, or international) could also affect citation or readership counts. Papers may get more citations due to their larger audience in a given country, or because of their different number of collaborators from different countries (Thelwall & Sud, 2016). These factors could increase visibility of research across different nations and thus associate with higher citation or readership counts. Results of cross-country analysis of Mendeley readers and authors of publications showed that Mendeley users mostly tend to select papers from their own countries, because they are more familiar with authors or fields from their own countries (Thelwall & Maflahi, 2015). High-impact authors, institutes, and countries involved in papers, or high-impact journals in which papers are published could also attract more citations. For instance, Journal Impact Factor (JIF) and international teamwork associate with increased citation impact in certain fields, including Biology and Biochemistry, Chemistry, and Social sciences (Didegah & Thelwall, 2013). However, international collaboration does not necessarily associate with high citation or readership counts. For instance, although the number of institutes and international collaborations associate with decreased citation impact in the field of Biochemistry, only collaboration with high-impact countries, such as the U.S., associates with increased citation and readership counts in this field (Sud & Thelwall, 2016).

In most fields, with the exception of Biomedical and health sciences, a negative relationship between number of pages and citations and Mendeley readership counts exists. This is in contrast, however, to other studies that showed that papers with more pages tend to attract more citations (Haustein et al., 2015; Ajiferuke & Famoye, 2015; Davis et al., 2011). This contradiction could be due to different sets of publications or choice of statistical tests (simultaneous vs. separate assessment of variables using regression or correlation analyses[7]) used in these studies. Longer papers may present more results and include figures, tables, references, etc., which may indicate a more in-depth analysis or more analytical research. It seems that this is an important feature for authors and Mendeley users in Biomedical and health sciences while in all other fields shorter papers are more attracted to them. It may be the case that the preference for condensed research or easy to read papers are prevalent by both authors and Mendeley users in these fields.

The relationship between title length with citations and readership varies across fields. For instance, in Biomedical and health sciences, Life and earth sciences, and Natural sciences and engineering, readership decreases with longer title length, while citations increase. This may reflect that, opposite to citations, technical titles are not attracted by Mendeley readers. These results are also supported by (Didegah et al., in press) who also found that title length associates with decreased Mendeley reader counts. Similarly, the probability of papers to be tweeted decreases with the lengths of titles (Haustein et al., 2015). Another study found that papers with more characters in their titles received more Mendeley readers than tweets compared to those with fewer characters (Xu, Khalili, & Deng, 2017). However, in Social sciences and humanities longer titles associate with more citation and readership counts as papers with longer titles may inform more specific, detailed, or scientific information in this field. Previous studies have shown that preference for this characteristic varies across different fields, and hence mixed results have been reported. For instance, in fields such as Psychology (Subotic & Mukherjee, 2014; Haslam et al., 2008) and Marketing (Stremersch et al., 2015), longer titles associate with fewer citations, while in Medical fields they associate with more citations (Jacques & Sebire, 2010). In PLOS journals, papers with longer titles receive fewer citations and are downloaded less frequently than those with longer titles. However, these results vary for papers with different title characteristics (having colons, question marks, etc.) (Jamali & Nikzad, 2011).

In the results presented above, only a few differences were found between citation and readership counts regarding the relation of number of institutes and title lengths in certain fields, while the results for all other variables were identical to citation impact. We therefore conclude that factors influencing readership and citation counts are broadly similar across all fields. It seems that though differences exist, citation and readership counts are affected by similar variables regardless of whether authors or Mendeley users have the same or different motivations to cite or save documents. This is to be expected due to the similarities between Mendeley users (mostly academics) and authors. From a more conceptual point of view, the broadly similar relationship between readership and citations can be explained by the use of Mendeley in a pre-citation context (Haustein, Bowman, & Costas, 2016) and by its stronger use among academic users (Zahedi, Costas, & Wouters, 2014). A survey of Mendeley users

[7] In the regression analysis the effect of one variable to another is assessed while controlling for the effect of other variables in the model (simultaneous assessment of variables). In contrast, only the relation between one variable with another variable are assessed by the correlation analysis regardless of controlling for the effect of other variables in the model (separate assessment of variables).

showed that this tool is used mostly by students, and that most Mendeley users use the tool to cite literature in their publications, followed by professional purposes (i.e., updating for job), teaching, and educational activities (Mohammadi, Thelwall, & Kousha, 2016).

However, both citation and readership counts are influenced by a range of motivations depending on the preference of authors and users when citing and saving documents. The act of saving documents in Mendeley can be also linked to motivations different from those for citing[8], for example reading for self-awareness, teaching, curiosity, or other professional needs (Haustein et al., 2015). This explains for example why some document types are substantially saved in Mendeley although they are not particularly cited later on (e.g., editorial materials, letters, book reviews, meeting abstracts, or news items). Therefore, it can be argued that although citing and saving are related activities, they also have some fundamental differences. Hence, Mendeley users do not always "adhere to the same norms as citations" when saving a document (Haustein et al., 2015). Moreover, both citation and readership impact could be influenced by some random (i.e., perfunctory) effects (Waltman, van Eck, & Wouters, 2013) that contribute to reducing the reliability of both readership and citation counts. Overall, the existence of these different motivations for saving documents in Mendeley and citing them suggests that one should not fully associate readership with citation counts. However, understanding all possible reasons for an author or Mendeley user to prefer some documents to others when citing and/or saving documents requires a more qualitative analysis; which is beyond the scope of the present study. Moreover, all the above results could be influenced by properties related to other variables (for instance author's gender, country, and reputation or impact; cited reference's impact and recency, etc. as discussed by Stremersch et al., 2015; Bornmann et al., 2012) and by specific norms and peculiarities of each field. Furthermore, the relationships between documents characteristics and citation and readership counts found in this study don't imply causation.

Finally, the limitations of this study need to be acknowledged. One limitation is that the variables analyzed are restricted to the same variables as studied by Haustein et al. (2015) in order to have comparable values to other altmetric sources. However, there could be other cultural, technological, economical, and political variables that could have an important effect on citation and readership counts. The extent to which researchers decide to use online platforms (like Mendeley) in their research workflow is influenced by many other external factors (Jamali et al., 2015; Gruzd, Staves, Wilk, 2012; Rowlands et al., 2011), such as different levels of familiarity with online platforms, or the age of researchers (e.g., younger researchers use social media more often than older researchers (Bolton et al., 2013; Bowman, 2015; Tenopir et al., 2015; Nicholas et al., 2015). For example, the different use and adoption of social media by different communities (as discussed by Bolton et al., 2013) could play an important role in why for some areas and some document types have a stronger (or weaker) relationship with Mendeley readership. Moreover, other document characteristics- such as openness (e.g., being open access), language of the publication, database coverage, availability of DOIs, or other different identifiers for the same document, the lack of complete

[8] For a review on citation behaviour see Bornmann & Daniel (2008).

metadata, etc.- may directly influence the rate of citation and readership received by a paper (Thelwall, 2015; Zahedi, Bowman, & Haustein, 2014). From a research evaluation perspective, one also needs to consider the underlying biases of Mendeley. Hence, there are age biases as Mendeley is used by younger researchers (Haustein & Larivière, 2014). There are also geographical biases in the coverage of publications, with some countries strongly covered and represented on Mendeley (Thelwall & Maflahi, 2015). All these biases call for caution in the interpretation of factors that could influence readership counts, particularly regarding the practical application of this type of indicator in any applied context.

4.5. Acknowledgements

The authors would like to thank the anonymous reviewers for their comments on this paper. Special thanks to Timothy Bowman from Wayne State University for his comments on the paper. The idea for this study came out when Zohreh Zahedi was a visiting scholar at the Canada Research Chair on the Transformations of Scholarly Communication (Université de Montréal), during Summer 2014, supported partly by the grant awarded by the Leiden University Fund (LUF #4509) and partly by funding from the Iranian Ministry of Science, Research, & Technology scholarship program (MSRT grant #89100156).

4.6. References

Ajiferuke, I., & Famoye, F. (2015). Modelling count response variables in informetric studies : Comparison among count, linear, and lognormal regression models. *Journal of Informetrics*, *9*(3), 499–513. https://doi.org/10.1016/j.joi.2015.05.001

Alperin, J. P. (2015). Geographic variation in social media metrics: an analysis of Latin American journal articles. *Aslib Journal of Information Management*, *67*(3), 289–304. https://doi.org/10.1108/AJIM-12-2014-0176

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly*, *25*, 152–179. https://doi.org/10.1016/j.leaqua.2013.10.014

Bolton, R. N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., … Solnet, D. (2013). Understanding Generation Y and their use of social media: a review and research agenda. *Journal of Service Management*, *24*(3), 245–267. https://doi.org/10.1108/09564231311326987

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. https://doi.org/10.1108/00220410810844150

Bornmann, L., & Haunschild, R. (2016). Normalization of Mendeley reader impact on the reader- and paper-side : A comparison of the mean discipline normalized reader score (MDNRS) with the mean normalized reader score (MNRS) and bare reader counts. *Journal of Informetrics*, *10*(3), 776–788. https://doi.org/10.1016/j.joi.2016.04.015

Bornmann, L., & Leydesdorff, L. (2015). Does quality and content matter for citedness ? A comparison with para-textual factors and over time. *Journal of Informetrics*, *9*(3), 419–429. https://doi.org/10.1016/j.joi.2015.03.001

Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, *8*(1), 175–180. https://doi.org/10.1016/j.joi.2013.11.005

Bornmann, L., Schier, H., Marx, W., & Daniel, H. D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*, *6*(1), 11–18. https://doi.org/10.1016/j.joi.2011.08.004

Bourdieu, P. (1998). *On television*. New York: New Press.

Bowman, T. D. (2015, July). *Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter* (Dissertation). Indiana University, Bloomington, IN, USA.

Boyack, K. W., & Klavans, R. (2005). Predicting the Importance of Current Papers. In *In P. Ingwersen , & B. Larsen (Eds.), Proceedings of the 10th meeting of the International Society for Scientometrics and Informetrics (ISSI 2005), Stockholm, Sweden: Karolinska University Press.* (pp. 335–342).

Buter, R. K., & van Raan, A. F. J. (2011). Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. *Journal of Informetrics*, *5*(4), 608–617. https://doi.org/10.1016/j.joi.2011.05.008

Costas, R., Haustein, S., Zahedi, Z., & Larivière, V. (2016). Exploring paths for the normalization of altmetrics: applying the Characteristic Scores and Scales. *The 2016 Altmetrics Workshop,* 27 September 2016, Bucharest, Romania.

Costas, R., Zahedi, Z., & Wouters, P. (2015a). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019.

https://doi.org/10.1002/asi.23309
Costas, R., Zahedi, Z., & Wouters, P. (2015b). The thematic orientation of publications mentioned on social media Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, *67*(3), 260–288. https://doi.org/10.1108/AJIM-12-2014-0173
Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *25*(7), 2129–34. https://doi.org/10.1096/fj.11-183988
Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*(4), 861–873. https://doi.org/10.1016/j.joi.2013.08.006
Didegah, F., Bowman, T. D., & Holmberg, K. (in press). On the differences between citations and altmetrics: An investigation of factors driving altmetrics vs. citations for Finnish articles. *arXiv preprint* arXiv:1710.08594.
Fox, C. W., & Boris, C. E. P. (2016). Citations increase with manuscript length , author number , and references cited in ecology journals, (August), 1–10. https://doi.org/10.1002/ece3.2505
Gruzd, A., Staves, K., & Wilk, A. (2012). Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model. *Computers in Human Behavior*, *28*(6), 2340–2350. https://doi.org/10.1016/j.chb.2012.07.004
Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, *28*(4), 39–47.
Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, *76*(1), 169–185. https://doi.org/10.1007/s11192-007-1892-8
Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting Altmetrics: Viewing Acts on Social Media through the Lens of Citation and Social Theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication* (pp. 372–406). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110308464-022
Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE*, *10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495
Haustein, S., & Larivière, V. (2014). Mendeley as a Source of Readership by Students and Postdocs ? Evaluating Article Usage by Academic Status. In *35th International Association of Technological University Libraries (IATUL),* 2-5 June 2014, Aalto University, Helsinki, Finland.
Jacques, T. S., & Sebire, N. J. (2010). *The impact of article titles on citation hits: an analysis of general and specialist medical journals. JRSM short reports, 1(1)*.
Jamali, H. R., Nicholas, D., & Herman, E. (2016). Scholarly reputation in the digital age and the role of emerging platforms and mechanisms, *25*(November 2015), 37–49. https://doi.org/10.1093/reseval/rvv032
Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, *88*(2), 653–661. https://doi.org/10.1007/s11192-011-0412-z

Larivière, V., & Gingras, Y. (2011). Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation. *Journal of Informetrics*, *5*(3), 392–399. https://doi.org/10.1016/j.joi.2011.02.001

Leimu, R., & Koricheva, J. (2005). Does Scientific Collaboration Increase the Impact of Ecological Articles? *BioScience*, *55*(5), 438. https://doi.org/10.1641/0006-3568(2005)055[0438:DSCITI]2.0.CO;2

MacRoberts, M. H. and MacRoberts, B. R. (2017), The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23970

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, *101*(1), 337–356. https://doi.org/10.1007/s11192-014-1345-0

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

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. https://doi.org/10.1002/asi.23477

Nair, L. B., & Gibbert, M. (2016). What makes a "good" title and (how) does it matter for citations? A review and general model of article title attributes in management science. *Scientometrics*, *107*(3), 1331–1359. https://doi.org/10.1007/s11192-016-1937-y

Nicholas, D., Jamali, H. R., Watkinson, A., Herman, E., Tenopir, C., Volentine, R., … Levine, K. (2015). Do Younger Researchers Assess Trustworthiness Differently when Deciding what to Read and Cite and where to Publish? *International Journal of Knowledge Content Development {&} Technology*, *5*(2), 45–63. https://doi.org/10.5865/IJKCT.2015.5.2.045

Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, *66*(4), 739–764. https://doi.org/10.1002/asi.23209

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS (" Leiden") evaluations of research performance. *Journal of Informetrics*, *4*(3), 423–430. https://doi.org/10.1016/j.joi.2010.02.003

Rovai, A. P., Baker, J. D., and Ponton, M. K. (2013). *Social Science Research Design and Statistics: A Practitioner's Guide to Research Methods and IBM SPSS*. VA: Water tree Press LLC.

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow, *Learned Publishing,* 24(3), 183–195. https://doi.org/10.1087/20110306

Shema, H., Bar-Ilan, J., & Thelwall, M. (2015). How is research blogged? A content analysis approach. *Journal of the Association for Information Science and Technology*, *66*(6), 1136–1149. https://doi.org/10.1002/asi.23239

Stremersch, S., Camacho, N., Vanneste, S., & Verniers, I. (2015). Unraveling scientific impact: Citation types in marketing journals. *International Journal of Research in Marketing*, *32*(1), 64–77. https://doi.org/10.1016/j.ijresmar.2014.09.004

Subotic, S., & Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, *40*(1), 115–124.

Sud, P., & Thelwall, M. (2016). Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology*, 67(8), 1849–1857.

https://doi.org/10.1002/asi.23515
Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (n.d.). Scholarly use of social media and altmetrics: a review of the literature. *Journal of the Association for Information Science and Technology*, 68(9). https://doi.org/10.1002/asi.23833
Tabachnick, B.G. and Fidell, L.S. (2007), *Using Multivariate Statistics* (5th ed.). New York: Allyn and Bacon.
Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3), 1195–1225. https://doi.org/10.1007/s11192-016-1889-2
Tenopir, C., King, D. W., Christian, L., & Volentine, R. (2015). Scholarly article seeking, reading, and use: A continuing evolution from print to electronic in the sciences and social sciences. *Learned Publishing*, *28*(2), 93–105. https://doi.org/10.1087/20150203
Thelwall, M. (2017). 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. https://doi.org/10.1177/0961000615594867
Thelwall, M. (2016). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *Journal of Informetrics*, 336–346. http://doi.org/ 10.1016/j.joi.2015.12.007
Thelwall, M. (2017). Avoiding obscure topics and generalising findings produces higher impact research. *Scientometrics*, *110*(1), 307–320. https://doi.org/10.1007/s11192-016-2159-z
Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields?. *Scientometrics*, 113 (3): 1721–1731. https://doi.org/10.1007/s11192-017-2557-x
Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, *8*(5), e64841. https://doi.org/10.1371/journal.pone.0064841
Thelwall, M., & Maflahi, N. (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*, *66*(6), 1124–1135. https://doi.org/10.1002/asi.23252
Thelwall, M., & Sud, P. (2015). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology*, *14*(4), n/a-n/a. https://doi.org/10.1002/asi.23559
Thelwall, M., & Sud, P. (2016). National, disciplinary and temporal variations in the extent to which articles with more authors have more impact: Evidence from a geometric field normalised citation indicator. *Journal of Informetrics*, *10*(1), 48–61. https://doi.org/10.1016/j.joi.2015.11.007
Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, *8*(4), 963–971. http://doi.org/10.1016/j.joi.2014.09.011
Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, *512*(7513), 126–129. https://doi.org/10.1038/512126a
Vieira, E. S., & Gomes, J. A. N. F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13. https://doi.org/10.1016/j.joi.2009.06.002
Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, *87*(3), 467–481. https://doi.org/10.1007/s11192-011-0354-5
Waltman, L., van Eck, N. J., & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of Informetrics*, *7*(3), 635–641.

https://doi.org/10.1016/j.joi.2013.04.001
Williams, P., Stevenson, I., Nicholas, D., Watkinson, A., & Rowlands, I. (2009). The role and future of the monograph in arts and humanities research. *Aslib Proceedings*, *61*(1), 67–82. http://doi.org/10.1108/00012530910932294
Xu, T., Khalili, S., & Deng, C. (2017). A study on factors related to readership of scientific articles. *STEM Fellowship Journal*, 3(2), 1-15.
Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does interdisciplinary research lead to higher citation impact? the different effect of proximal and distal interdisciplinarity. *PLoS ONE*, *10*(8), 1–21. https://doi.org/10.1371/journal.pone.0135095
Zahedi, Z., Bowman, T., & Haustein, S. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. In *Metrics 2014, ASIS&T Workshop on Informetric and Scientometric Research*, Seattle, US.
Zahedi, Z., Costas, R., Larivière, V., & Haustein, S. (2016). What makes papers visible on social media ? An analysis of various document characteristics. In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *21th International Conference in Science & Technology Indicators (STI), 13-16, September, 2016, Valencia, Spain.* (pp. 13–16).
Zahedi, Z., Costas, R., & Wouters, P. (2014). Assessing the Impact of Publications Saved by Mendeley Users: Is There Any Different Pattern Among Users? In *35th International Association of Technological University Libraries (IATUL),* 2-5 June 2014, Aalto University, Helsinki, Finland.
Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, 68 (10): 2511-2521. https://doi.org/10.1002/asi.23883

CHAPTER 5

Mendeley Readership as a Filtering Tool to Identify Highly Cited Publications[1]

[1] This chapter is based on the peer reviewed version of the following article which has been published in final form at http://doi.org/10.1002/asi.23883. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions:

Zahedi, Z., Costas, R., and Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. Journal of the Association for Information Science and Technology, 68 (10): 2511-2521. http://doi.org/10.1002/asi.23883

Abstract

This chapter presents a large scale analysis of the distribution and presence of Mendeley readership scores over time and across disciplines. We study whether Mendeley readership scores (RS) can identify highly cited publications more effectively than journal citation scores (JCS). Web of Science (WoS) publications with DOIs published during the period 2004-2013 and across 5 major scientific fields have been analyzed. The main result of this study shows that readership scores are more effective (in terms of precision/recall values) than journal citation scores to identify highly cited publications across all fields of science and publication years. The findings also show 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 scores 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 vs. 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.

Keywords

Mendeley readership scores; Journal citation scores; highly cited publications; precision-recall analysis

5.1. Introduction and background

Scholars use social media tools for different purposes, for example to collaboratively distribute scientific information, share knowledge and ideas, and communicate with their peers (Gruzd, Staves, & Wilk, 2012). Among the different altmetric sources, Mendeley is one of the most important online reference managers with more than 4 million users worldwide[1], and is especially popular among students and postdocs (Zahedi, Costas, & Wouters, 2014b; Haustein & Larivière, 2014). Mendeley exhibits a high coverage of scientific publications, with coverage values higher than 60% or even 80% for WoS publications depending on the field (Costas, Zahedi, & Wouters, 2015b).

5.1.1. Meaning of Mendeley readership

Mendeley collects usage statistics per document as they are added by the different users to their private libraries. These statistics are commonly known as “readership statistics”, although in reality the metrics don’t necessarily reflect the actual ‘reading activity’ by Mendeley users. For example, scholars do not necessarily always ‘read’ the scholarly outputs that they save in Mendeley (Mohammadi, Thelwall, & Kousha, 2015). Thus the actual meaning of “readership” in Mendeley is not fully known yet and this introduces a conceptual constraint on the actual value that the act of “saving” a document in Mendeley may have. Moreover, not all scholars are familiar with Mendeley; instead they may use other reference management tools in their scholarly process of reading and referencing papers (or none at all). Therefore, the usefulness of Mendeley readership strongly depends on the coverage and presence of users from different disciplines, countries, academic statuses, ages, etc. Another important issue is that Mendeley does not provide any information on the timestamp (date) when a given document has been added by a user to her/his library[2]. Therefore, important information on the patterns of readership scores accumulation over time for the saved publications is still lacking, making the adequate study of readership history patterns impossible.

5.1.2. Characteristics of Mendeley as a scientometric tool

Previous studies have shown moderate correlations between readership and citation scores (see Zahedi, Costas, & Wouters, 2014a; Haustein et al., 2014b; Thelwall & Wilson, 2015). The correlations between Mendeley readership and citation scores are higher than the correlations between citations and other altmetric indicators (Thelwall et al, 2013; Costas, Zahedi, & Wouters, 2015a), thus a stronger similarity between these two metrics in comparison to other altmetric sources can be assumed. Furthermore, publications with more Mendeley readership scores tend to have higher number of citations and are published in

[1] http://blog.mendeley.com/elsevier/mendeley-and-elsevier-2-years-on/

[2] Users in Mendeley can view only the historical overview of readership (the last 12 months) of their own documents saved in Mendeley (this information is not yet available via API and for all the documents saved in Mendeley by all users).

journals of higher impact compared to those with less or without any readership (Zahedi, Costas, & Wouters, 2014a). All these results suggest that Mendeley can be a relevant tool for scientometric purposes, and for example, suggestions of normalization of the number of readership by discipline have already been proposed (Haunschild & Bornmann, 2016). Some other important features of Mendeley are that these readership statistics include data about the academic's status, disciplines and countries of the Mendeley users. This information on the academic's disciplinary and geographic background of the different users helps to better understand the saving patterns of scientific publications by different groups of users (Haunschild, & Bornmann, 2015; Haunschild, Bornmann & Leydesdorff, 2015; Thelwall & Maflahi, 2015). Another important characteristic of this tool is that readership data tend to be collected and made available before citation is recorded by any citation database. Thus, Mendeley readership scores can be seen as evidence of 'early' impact of scientific publications (Maflahi & Thelwall, 2016). However, as mentioned before, due to the lack of historical information reported by Mendeley regarding the date and time at which readership happened, it is not possible, at this time, to perform reliable analyses regarding the prediction of future citations using Mendeley readership scores.

5.1.3. Identification of highly cited publications

Studying highly cited publications and the factors influencing them is an important topic in the scientometric literature (Ivanović & Ho, 2014; Aksnes, 2003). Although being highly cited does not always truly reflect the higher research quality of publications (Waltman, Van Eck, & Wouters, 2013), high citedness can be a characteristic sign of relevant or even potential 'breakthrough' papers (Schneider & Costas, 2014) as well as an indicator of scientific excellence (Bornmann, 2014) and the share of such highly cited papers is considered as a relevant indicator in research evaluation in a large number of fields (Abramo et al., 2015; Tijssen et al., 2002). Therefore the identification of highly cited publications can be considered as a critical element in bibliometric research as well as research evaluation. The use of journal level impact indicator in order to capture the "quality" of individual scientific publications has been widely criticized in the literature (Adler, Ewing, & Taylor, 2008). They have been observed to have weak correlations with citations at the publication level and they are not well representative of individual article impact (Seglen, 1997; Larivière et al, 2016) and can be influenced by highly cited publications (Seglen, 1992). As a reaction to this, some initiatives such as DORA[3] and the Leiden manifesto[4] have warned against the misuse of journal-based indicators in the evaluation of publications and individuals. On the other hand, high journal impact indicators may indicate a higher probability that some publications in the journal will attract large numbers of citations (although we do not know beforehand which ones will be the most highly cited). In addition, authors tend to see publication in high impact journals as a strong performance in itself since these journals are often highly selective. For instance,

[3] DORA: San Francisco Declaration on Research Assessment: www.ascb.org/dora/
[4] Leiden Manifesto for Research Metrics: www.leidenmanifesto.org/

Biomedical researchers in the Netherlands perceive the quality and novelty of papers by the impact of journal (namely JIF) in which these papers are published (Rushforth, & Rijcke, 2015). The combined use of journal and publication level impact indicators (so called "composite indicator") has been proposed for evaluating recent publications (Levitt & Thelwall, 2011; Stern, 2014). It has also been shown that using geometric vs. arithmetic mean in calculation of journal impact factor helps to reduce the influence of highly cited publications on its correlation with individual publications (Thelwall& Fairclough, 2015). In a similar line, using journal level metrics in evaluating research has been seen as a relevant practice in some countries (e.g., in Spain, Jiménez-Contreras et al, 2003). In addition, there have been discussions about the potential relevance of journal-based indicators as tools for the analysis of researchers and, particularly, for the potential filtering and selection of academic papers for reading (cf. Waltman, 2016). In this paper we follow up on the latter argument (i.e., the relevance of using journal-based indicators to filter highly cited publications) in contrast to Mendeley readership.

5.1.4. Justification and aim of this study

It will be clear that if alternative metrics do not improve the ability of filtering highly cited publications of journal indicators, they don't really pose a true advantage over currently existing measures of impact (e.g., the Journal Impact Factor) for this purpose. The study of the ability of altmetric indicators to identify highly cited publications in comparison with journal-based indicators has shown that journal-based indicators have both a stronger correlation with citations as well as stronger filtering power to identify highly cited publications than for example F1000 recommendations, tweets, blogs as well as other altmetric indicators (Waltman & Costas, 2014; Costas, Zahedi, & Wouters, 2015a). These results reinforce the idea that the above mentioned altmetric indicators do not introduce any advantage over journal-based indicators to identify highly cited publications. Mendeley hasn't been thoroughly studied yet from this perspective. If Mendeley would offer a better filtering solution for identifying highly cited publications than journal indicators, it could be argued that "at least" Mendeley readership represent a true alternative to journal indicators when screening for relevant publications. Actually, in a preliminary study it has been reported that readership scores are more effective at identifying highly cited publications than journal citation scores for the 2011 Web of Science publications (Zahedi, Costas & Wouters, 2015). Thus, as mentioned before, in contrast to other altmetric indicators, this finding indicated for the first time that Mendeley readership scores could represent a valuable tool as an alternative to journal indicators for a more effective filtering of highly cited publications. Due to the relevance of such a result, and since the previous study was limited to only one publication year, in this study we aim to extensively test whether this pattern is also present in data sets with longer publication and citation windows as well as from different scientific disciplines. Thus, the main aim of this paper is to explore the relationship between Mendeley readership and journal citation scores, particularly focusing on whether Mendeley readership scores are able to identify highly cited publications more effectively than journal-based impact indicators.

5.2. Data and Methodology

This study is based on a dataset of 9,152,360 (77.5%)[5] Web of Science (WoS) publications (articles and reviews) with Digital Object Identifiers (DOI) from the years 2004-2013. The readership data from Mendeley were extracted via Mendeley REST API on February 9, 2015. 86.5% (7,917,494) of all papers have at least one Mendeley readership while 13.5% (1,234,866) of them don't have any. A variable citation window (i.e., citations from 2004 until the end of the year 2014) has been considered for calculating the citation scores. Self-citations have been included in the citation scores in order to keep the same approach for citation and readership data, since it is not possible to calculate something like "self-readership" in Mendeley. Moreover, due to the lack of information on the date of documents added to users' libraries in Mendeley (the date of readership), it is not possible to exactly establish the same citation and readership windows. Hence, we consider the sum of all readership data until 9 February 2015 as the total readership score. The journal citation score and top 10% most highly cited publications[6] in the period 2004-2014 have been calculated for each publication. Only document types 'article' and 'review' were considered. The fields to which publications belong were determined according to the five major fields of science in the 2013 Leiden Ranking classification[7]. The following indicators have been calculated for the different analysis using the CWTS in-house database:

P: *total number of publications (articles and reviews).*

Total Citation Score (TCS)*: sum of all the citation scores received by the publications in the period of 2004-2014.*

Total Readership Score (TRS)*: sum of all Mendeley readership scores (RS) received by the publications until February 2015.*

Mean Citation Score (MCS): *average number of WoS citation scores per publication.*

Mean Readership Score (MRS): *average number of Mendeley readership scores per publication.*

Journal Citation Score (JCS)[8]: *average number of WoS citations received by all publications in a journal in a period of 2004-2014.*

---

[5] 77.5% of all WoS articles and reviews from the years 2004-2013 have a DOI.

[6] Top 10% publications are publications that belong to the top 10% quartile of the most cited publications in their fields (i.e., Web of Science Subject Categories) and publication years. We have followed the methodology by Waltman & Schreiber (2013) for the calculation of percentile based indicators, although in this case proportionally assigned publications to the top 10 percentile have been considered as fully top 10% highly cited publications. Also all articles and reviews in the WoS database (i.e., including papers without DOIs and not covered by Mendeley) are considered for the determination of the top 10% highly cited publications.

[7] www.leidenranking.com/ranking/2013

[8] JCS calculation is based on all outputs of the journals (i.e., regardless of having or not DOIs and even if not all of them are covered by Mendeley).

The distribution of the above indicators over time and across their subject fields has been investigated. This has been done in order to provide a general overview of the data and to identify any relevant pattern regarding the density of readership in comparison to citation scores across fields and publication years. A precision-recall analysis (Harman, 2011) has been performed in order to evaluate the ability of readership scores and journal citation scores to identify highly cited publications. In information retrieval, precision is the proportion of retrieved documents that are relevant, while recall is the proportion of relevant documents that are retrieved. Accordingly, in this study for a given selection of publications, precision is the ratio (%) of highly cited publications divided by the total number of publications in the selection, and recall is the ratio (%) of highly cited publications in the selection divided by the total number of highly cited publications (Waltman & Costas, 2014). All the 'top 10%' highly cited publications in the sample have been identified. Then, publications have been ranked by their individual readership scores in a descending order (ties have been sorted randomly) and the precision-recall analysis has been performed. The same process was performed using the journal citation score of the journal of each individual publication. Thus two precision-recall analyses have been produced, one for the readership scores and another one for the journal citation scores. Finally, the values have been plotted, where the *x* axis represents the 'Recall' and *y* axis represents the 'Precision' values. The precision-recall analysis has been done both across publication years (from 2003-2014) and also across subject fields (based on the 5 major fields of science in the 2013 Leiden Ranking classification) of the publications. The precision-recall curves provide visual representations of how precision values correspond with their recall values.

5.3. Results

5.3.1. General distribution of citation and readership scores over time

Table 1 shows the descriptive statistics for the entire publication set used in this study. In general, the average number of citation per publication (MCS) is higher than the average number of readership score per publication (MRS), which means that on average all publications received more WoS citation sores than Mendeley readership scores. The table also shows that the coverage of publications with at least one Mendeley readership is increasing from 2004 to 2009 with a decrease for the most recent years (from 2010 until 2013).

Table1. General distributions of MRS and MCS indicators of the WoS publications across publication years 2004-2013.

Pub year	P	Cov	%	TRS	MRS	TCS	MCS
All years	**9,152,360**	**7,917,494**	**86.51**	**102,051,962**	**11.15**	**132,246,959**	**14.44**
2004	540,924	458,114	84.69	6,129,245	11.33	15,724,035	29.07
2005	618,976	531,409	85.85	7,452,051	12.04	16,706,508	26.99
2006	713,864	615,637	86.24	8,697,103	12.18	16,990,568	23.80
2007	788,533	682,704	86.58	9,801,854	12.43	16,669,281	21.14
2008	872,572	768,813	88.11	11,252,702	12.90	16,084,499	18.43
2009	962,262	857,585	89.12	12,547,495	13.04	15,106,704	15.70
2010	1,026,541	913,414	88.98	13,260,840	12.92	13,026,893	12.69
2011	1,120,212	987,479	88.15	12,909,807	11.52	10,504,765	9.38
2012	1,206,707	1,030,886	85.43	11,217,458	9.30	7,499,214	6.21
2013	1,301,769	1,071,453	82.31	8,783,407	6.75	3,934,492	3.02

Number of publications (P); Coverage (n. pubs) in Mendeley per publication year (Cov); Total Readership Score (TRS); Mean Readership Score (MRS); Total Citation Score (TCS); Mean Citation Score (MCS)

According to figure 1, MCS steadily decreases over these 10 years; while, on the other hand, MRS first follows a relatively stable pattern with a small increase from 2004 to 2009 and then shows a decrease from the year 2010 onwards, in which MRS is higher than MCS. The higher density of MRS over MCS for publications has also been observed in previous studies on Mendeley (Haustein & Larivière, 2014; Thelwall, 2015; Maflahi, & Thelwall, 2016; Zahedi, Costas, & Wouters, 2015; Costas, Zahedi, & Wouters, 2015a). This suggests that the more recent publications received on average more readership than citation scores. These results support the idea of a faster accumulation of Mendeley readership scores over publications in contrast to citation scores.

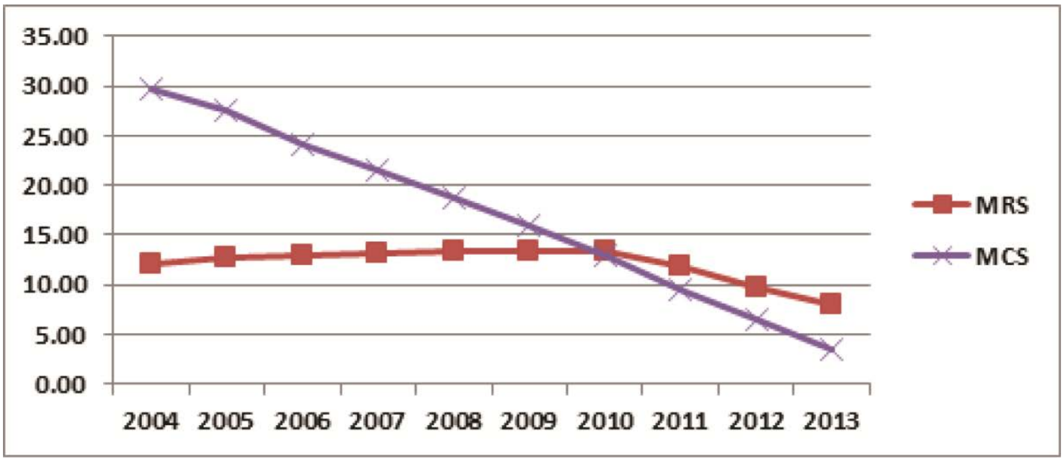


Figure1. Distributions of MRS and MCS indicators for the WoS publications overtime (x axis shows the publication years and y axis shows the mean scores of citation and readership).

5.3.2. General distribution of MCS and MRS indicators across fields

MRS and MCS indicators have been calculated for the publications based on their main disciplines in the 2013 Leiden Ranking (LR) classification. Table 2 presents the values of MCS and MRS for the 5 major LR fields of science. 'Biomedical and health sciences' is the biggest field with around 36% of all Mendeley-covered publications while 'Social sciences and humanities' is the smallest one in the dataset (7.6%). In terms of coverage of publications in Mendeley (i.e., based on publications (articles and reviews) with DOI), 93% of publications from 'Life & earth sciences' and 92% from 'Social sciences & humanities' have at least one reader in Mendeley, while just 77% of publications from 'Mathematics & computer science' have some readership in Mendeley. Also, the coverage of publications per LR fields with presence in Mendeley increases from 2004 to 2010 with a small decrease for the recent years (from 2011-2013) (see Appendix 3, Table 1 for all publication years). In terms of citation and readership frequency, 'Life and earth sciences' have on average the highest mean readership scores (MRS=18.64) followed by 'Social sciences and humanities' (MRS=18.14), whereas 'Biomedical and health sciences' have the highest mean citation scores (MCS=20.18). Publications from 'Mathematics and computer sciences' exhibit the smallest values both in terms of readership and citation scores (MRS=7.52 and MCS=8.0). In terms of citation and readership density, publications from Social sciences fields have a higher density of readership over citation scores. In contrast, publications from 'Biomedical and health sciences', although with the highest coverage in the sample, exhibit a lower readership density as compared to their citation density. These results are in line with previous analyses (Thelwall, 2015; Costas, Zahedi, & Wouters, 2015b) and indicate that, similar to citation, the readership density of publications varies per fields. Furthermore, large differences in the WoS database coverage across disciplines could affect the density of citations across subject fields, the same holds for the Mendeley database.

Table2. General distributions of MRS and MCS indicators for the WoS publications across LR fields.

Main fields	P	%	Cov	%	TRS	MRS	TCS	MCS
Biomedical & health sciences	3,340,837	35.90	3,033,467	90.80	45,468,376	13.60	67,437,722	20.18
Life & earth sciences	1,512,173	16.25	1,407,153	93.06	28,189,119	18.64	26,668,168	17.63
Mathematics & computer science	859,363	9.24	660,908	76.91	6,470,579	7.52	6,877,035	8.00
Natural sciences & engineering	2,878,982	30.94	2,409,731	83.70	23,641,874	8.21	43,656,107	15.16
Social sciences & humanities	714,142	7.67	659,754	92.38	12,956,645	18.14	7,346,205	10.28

Number of publications (P); Coverage (n. pubs) in Mendeley (Cov); Total Readership Score (TRS); Mean Readership Score (MRS); Total Citation Score (TCS); Mean Citation Score (MCS)

Comparing the distribution of citation and readership scores across fields of science, Figure 2 shows that for fields such as 'Social sciences and humanities' and 'Life and earth sciences' MRS values are higher than MCS values. These are also the fields with the highest coverage in Mendeley. This higher density of readership over citation is even bigger in the field of 'Social sciences and humanities' (MRS=18.14 vs. MCS=10.28). There are also variations in density of

MRS vs. MCS by the different LR fields across the different publication years (see Figure 1 in Appendix 3). Basically, for the oldest papers of all disciplines, MCS values are higher than MRS values, while MRS values are higher than MCS in all cases for the most recent years. The case of 'Social sciences and humanities' is different, as MRS outperforms MCS for all years except for the first year 2004 (see Figures 1 & 2 in Appendix 3) indicating that readership scores in this field have a much stronger density as compared to citations over a longer period of time. In order to further explore which subfields within the 'Social sciences and humanities' exhibit higher readership vs. citation densities, MRC and MCS values have been calculated for the individual WoS subject categories (Appendix 2.1). The results show that publications from fields such as Business, Psychology, Sociology, Social and behavioral sciences, Anthropology, Education and educational research and Linguistics are among the fields that have a higher readership density than citation density. Fields such as Chemistry, Oncology, Hematology, Physics, Medicine and Virology have the highest MCS values over MRS (see Appendixes 2.1 & 2.2). These results confirm the idea of important disciplinary differences in readership practices (see Thelwall & Sud, 2015; Costas, Zahedi, & Wouters, 2015b) in a very similar way as it has been observed for citation practices (see Waltman & Van Eck, 2013; Crespo, Li, & Ruiz–Castillo, 2013; Crespo et al., 2014). These differences highlight both different citing and reading practices across fields as well as the disciplinary differences in the coverage of citation and readership databases. Disciplinary differences have also been seen in the use of other academic social networking sites and other online reference managers. For example, Academia.edu is mostly used by academics from Social sciences and humanities in contrast to researchers from physical, health and life sciences, biology, medicine and material sciences with very low usage of this platform (Thelwall & Kousha, 2014; Mas-Bleda et al., 2014; Ortega, 2015). Similarly, CiteULike is known to be more popular among users from the biomedical domain (Hauff & Houben, 2011). Twitter has been shown to have a good coverage within the field of biomedicine (Haustein et al., 2014a). Twitter is also used by researchers from diverse disciplines such as biochemistry, astrophysics, chemoinformatics (field related to the use of computer techniques in chemistry) and digital humanities, and for different purposes such as scholarly communication, discussions, sharing links (e.g., in fields like economics, sociology and history of science) (Holmberg & Thelwall, 2014).

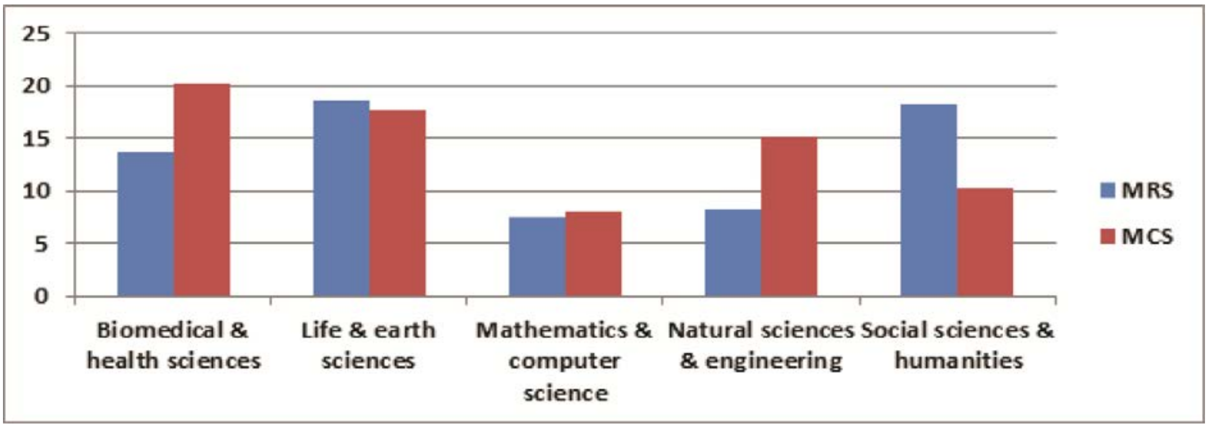


Figure2. Distribution of MRS and MCS indicators for the WoS publications across LR fields (*x* axis shows the fields and *y* axis shows the mean readership and mean citation scores).

Another study has observed the same variation between fields in the amount of citation and readership scores concluding that in some fields such as Ecology, Evolution, and Behavior and Systematics (based on Scopus subject categories), Mendeley scores are much higher than citations. Also, correlations between these Mendeley readership and citations have been found to have a decreasing trend for recent publications (2011 to 2014) (Thelwall & Sud, 2015). All in all, the coverage, language and any other biases related to the citation and readership databases could cause important limitations on research assessment and impact indicators, particularly in some fields with low coverage such as social sciences and humanities (Van Leeuwen et al, 2001). For instance, in the humanities, different information behaviours, dependency on print vs. online materials and database's low coverage of non-English publications influence the analysis of scholarly materials (Collins et al., 2012; Hammarfelt, 2014). As an alternative solution to any bias that a database may have, the combined use of citation databases has been proposed (Meho & Sugimoto, 2009). Further research should therefore focus on considering other databases and test if the elements discussed here also hold for them. For now, we still consider that an analysis based on the Web of Science has a strong relevance as this is one of the most common and used data sources for Scientometric and altmetric research.

5.3.3. Precision-Recall analysis of all publications in the sample

In order to test which of the two indicators (i.e., Mendeley readership scores or journal citation scores) is more effective to identify highly cited publications, precision-recall analyses have been performed across publication years and subject fields separately. Figure 3 shows the results of the general precision-recall analysis of RS over JCS for all the publications in the dataset over time. According to this figure, the RS (green line) performs better than JCS (blue line) in the whole spectrum of precision-recall in identifying the top 10% most cited publications in all publication years. The figure indicates that, for example, a recall of 0.5 (50%) corresponds with a precision of 0.45 (45%) for RS and a precision of 0.25 (25%) for JCS in the years 2004-2013. This means that if we want to select half of all highly cited publications in the dataset in each year, we have an error rate of 55% when the selection is made based on readership scores, and an error rate of 75% when the selection is made based on journal citation scores. Actually, error rate refers to the share of highly cited papers that cannot be identified by one of these two indicators (RS or JCS). In the precision-recall figure, by drawing a vertical line from the recall axis for example from the recall point of 0.5 (50%) crossing the RS and JCS lines, and drawing a horizontal line from there to the precision axis, it shows that the recall of 50% corresponds to a precision levels of 45% for RS and of 25% for JCS. This means that the error rates for RS is 100-45=55% and for JCS is 100-25=75%. The results of the figures are straightforward; the green line always outperforms the blue line in terms of precision in the whole spectrum of recall. Hence we can conclude that readership scores identify highly cited publications better than journal citation scores for all the publication years in our dataset. This is a very important result as it has not been observed before for other altmetric sources (cf. Costas, Zahedi, & Wouters, 2015a; Waltman & Costas, 2014).

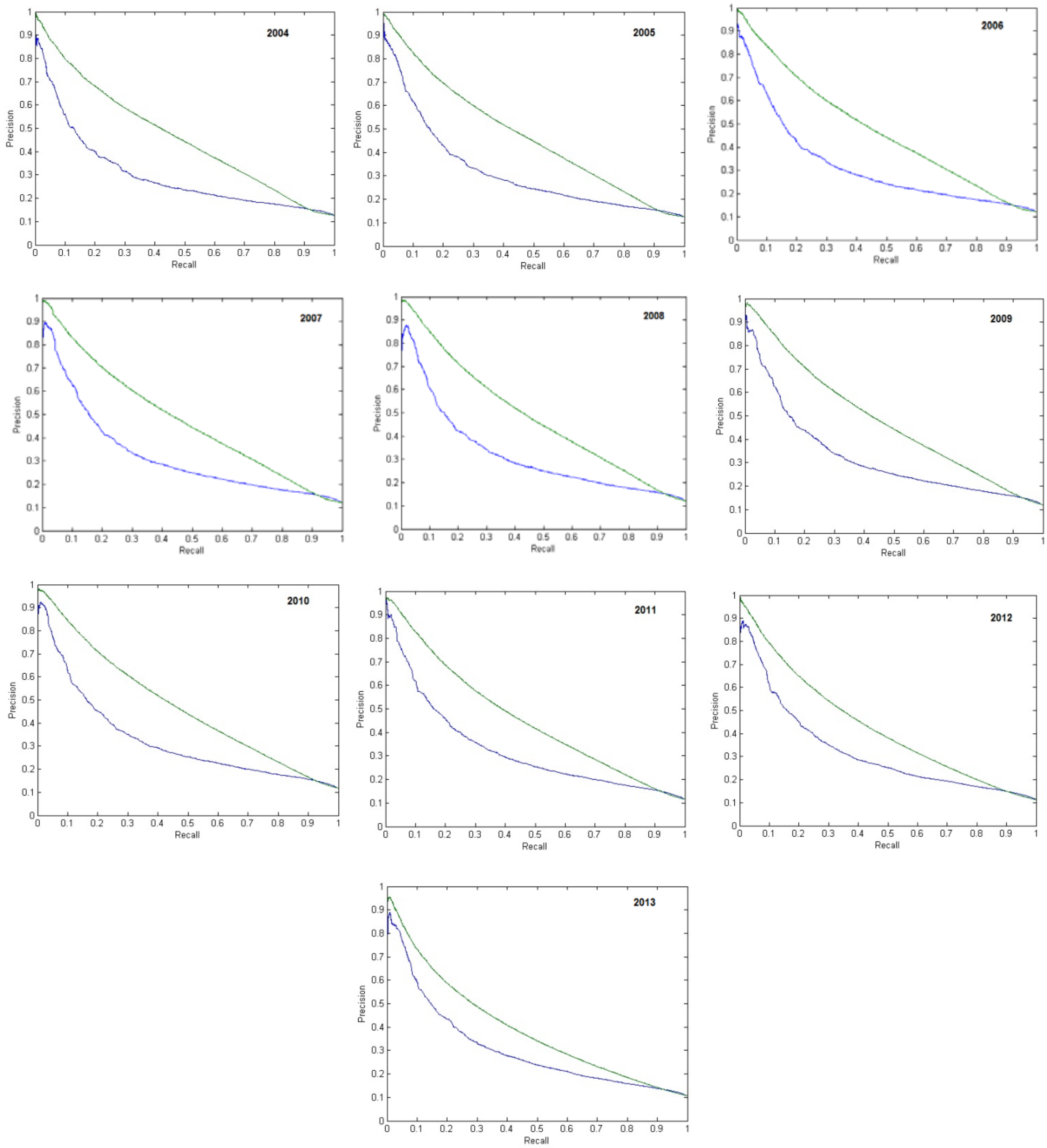


Figure 3. General Precision-recall curves for JCS (blue line) and RS (green line) for identifying top10% most highly cited WoS publications from the years 2003-2014 left to right (x axis represents the 'Recall' and y axis represents the 'Precision' values).

5.3.4. Precision-recall analysis of publications across their disciplines

In this section, the precision-recall analysis has been performed across disciplines. Results indicate that RS also outperforms JCS in identifying highly cited publications for all major fields of science. All the figures are similar, essentially resembling the general patterns in Figure 3. These results are in line with the result obtained for the 2011 WoS publications (Zahedi, Costas, & Wouters, 2015) confirming the better capacity of RS over JCS in identifying highly cited WoS publications for fields of science. Thus, this pattern can be considered to be robust both across disciplines and years (see also appendix 3). The only noticeable exception is the field of 'Mathematics & computer science'. In this field, JCS outperforms RS both in the lower (below 10%) and higher (above 80%) levels of recall. For example, for the publications from the years 2004 to 2009, RS outperforms JCS until the recall point of 0.5 (50%) while, for the most recent years (from 2010 onwards), there is a small advantage of JCS over RS particularly from the recall point of 0.5 onwards. A potential explanation for this exception is that this is the field with the lowest coverage of publications saved in Mendeley (76.91%) of the publications in this field are covered by Mendeley) as well as the field with the lowest density of both citation and readership scores compared to the other fields in the study. These lower coverage and density values could be more easily affected by all kinds of random effects coming from citation and also readership processes[9] (cf. Waltman, Van Eck, & Wouters, 2013), thus having a greater influence on the patterns observed for this discipline. According to the literature, the low citation rates of Mathematics and computer science compared to fields such as Chemistry or Physics can be also related to the specific publication and citation behaviours in these fields (Korevaar & Moed, 1996; Seglen 1997). For instance, scholars from fields like Mathematics and computer science are known to publish more in formats such as research reports and conference papers which are not included in citation databases such as Web of Science (Moed et al., 1985; Bornmann et al., 2008). Also, Mathematics is a discipline with a relatively low number of references per paper as compared to other disciplines (Vieira & Gomes, 2010; Glänzel & Schoepflin, 1999). This lower level of references per paper may explain the lower density of citations per paper in the field (i.e., there are fewer references (citations) pointing to other Mathematics papers) as well as lower numbers of Mendeley readership (i.e., Mendeley users from Mathematics would save fewer records in their Mendeley libraries). Other reasons for the low rates of readership also include the different orientation, uptake and use of Mendeley among scholars in this field. Users from Mathematics and computer science seem to be more oriented towards other reference managers such as BibSonomy (Hauff & Houben, 2011), which may support the idea that Mendeley is not the most popular online reference manager tools among the users of these fields.

All in all, the results of the precision-recall analysis highlight the importance and potential of Mendeley readership as a tool for research evaluation. This suggests that readership data can be used as a relevant tool in finding highly-cited publications. This result together with the fact that Mendeley readership are available both openly and also much earlier than citation as well as their potential in revealing an early impact of publications (Maflahi & Thelwall, 2016), put an emphasize on an additive value that readership data offer in case of its usage beside other impact indicators for any research evaluation and scientometrics purposes.

[9] The citation process is known for being "noisy" and influenced by multiple random factors that limit the relationships between citation and scientific impact (see Waltman, Van Eck, & Wouters, 2013). In a similar manner, we can argue that similar noisy factors can influence the relationship between the act of saving in Mendeley, citations and scientific impact.

5.4. Discussions and conclusions

This study presents a large-scale analysis of the distribution and presence of Mendeley readership scores over time and across disciplines. Precision-recall analysis has been used to test the ability of Mendeley readership scores to identify WoS highly cited publications, particularly in comparison with journal citation scores. Our results show that 86.5% of the publications in our dataset were covered in Mendeley with at least one reader. The coverage of publications with some Mendeley readership increased from 2004 to 2009 with a small decrease from 2010 onwards. Disciplinary differences have been found in terms of both citation and readership density. These differences of readership density could be explained by the different levels of awareness and adoption regarding the use of Mendeley in the scholarly practice of researchers (Ortega, 2015) or by the use of other reference managers such as BibSonomy or CiteULike (Hauff & Houben, 2011) by scholars from different fields. However, further research on this point is still needed.

The main conclusions of this study can be summarized as follows:

a) *Steady increase of Mendeley readership scores for the earliest publication years and decreasing pattern for the most recent ones.*

The average readership per publication steadily increases from 2004 until 2009, with a small decrease for the most recent years (i.e., 2010 onwards). This pattern is observable for all fields. These results are in line with those of Thelwall & Sud (2015) for a selection of Scopus thematic categories (including agriculture, business, decision science, pharmacy, and the Social sciences) and LIS journals (Maflahi & Thelwall, 2016).These authors found very similar steady increasing patterns for Mendeley readership for older years with a decrease for the most recent ones. A plausible explanation for this pattern (as opposed to the consistently higher average values of citations per paper for the older years) is that citations are events that can happen several times (i.e., a paper can be cited multiple times), but a paper can only be saved once by each Mendeley user. Thus, the maximum number of readership a paper can achieve is the total number of users in Mendeley, while the number of citations a paper can receive has basically no upper bound. Moreover, the removal of papers from Mendeley libraries[10] by users can contribute to explain the patterns observed for the older years. Thus, in order to maintain manageable libraries, Mendeley users could decide to remove the older and less useful publications from their reference managers. As a result, citations would always accumulatively increase over time as publications have more time to be cited, while the number of readership could actually decrease as users would remove older references from their libraries. Moreover, as pointed by Thelwall & Sud (2015), Mendeley was launched and

[10] According to William Gunn (Director of Scholarly Communications in Mendeley), "When a user deletes their account and all their documents, the readership of that document doesn't change, until the batch clustering process is re-run and the new number of metadata records is generated. The same applies when a user deletes a record from their library. In summary, the count of records can increase nearly instantaneously, but only decreases periodically" see:
www.niso.org/apps/group_public/view_comment.php?comment_id=632
www.niso.org/apps/group_public/view_comment.php?comment_id=610

became available in 2008 and consequently became popular afterwards. This may contribute to explain the increase in MRS values from 2008 to 2009.
Another possible reason for the decreasing pattern of readership for recent publications could be the delay between the publication of the paper and the time needed by the users to spot it and decide to save it in their libraries. In other words, the declining pattern for the most recent years is likely indicating some kind of delay in the accumulation of readership for the most recent publications. Finally, variations in the uptake of Mendeley across fields and the increasing popularity of other reference managers in some fields, as well as changes in the preferences of users in their reference manager choices (e.g., preferring Zotero over Mendeley) might have played an influence on the lower counts of Mendeley readership during the most recent years. However, the lack of reliable information on the uptake of reference manager among different types of users make difficult to determine the true importance of such as pattern. In any case, it is important to notice how even with this delay in the accumulation of readership, they accumulate faster than citations during the three most recent years.

b) *Higher density of Mendeley readership scores over citations for the most recent years and most disciplines*

Our results show that the density of Mendeley readership is higher than that of citations for the most recent years and for most of the disciplines. These results suggest the potential advantage of Mendeley readership over citations for the analysis of impact of the most recent publications and particularly in the field of Social sciences, which is also a field that traditionally is not well represented by citation databases (Nederhof, 2006). Thelwall & Sud (2015) suggested that the faster uptake and the stronger density of Mendeley reader counts for the most recent years could be seen as a good proxy for "early scientific impact" for articles from recent years and also for fields with higher levels of Mendeley use. However, our results also show that as time passes and more citations accumulate, they tend to outperform the values of readership (which tend to remain stable) after around 3-4 years, although again this varies across disciplines. For example, the readership advantage over citations lasts longer in the Social sciences than in the Natural sciences (see appendix 2.1). Maflahi & Thelwall (2016) found similar patterns for a set of LIS journals.
These results suggest that Mendeley readership scores can work as an important source to reflect evidence of "early impact" of scientific publications since, as shown in this study as well as in a previous analysis (Thelwall & Sud, 2015), readership occur and are available earlier than citations during the first years after publication. However, more research is necessary in order to better disentangle the true motivations of Mendeley users and differences between citations and Mendeley readership during the first years after publication of the articles.

c) *Higher filtering ability of highly cited papers by Mendeley readership scores in contrast to journal citation scores.*

The most important result of this study shows that Mendeley readership data can work as a relevant tool to identify highly cited publications in WoS. This finding is robust both across most major fields of science and publication years. In contrast, other altmetric indicators (e.g., F1000 recommendations, Twitter, blogs, etc.) have not been found to have such a property, particularly in their comparison with journal citation scores as a benchmark tool to identify highly cited publications. Based on these results, it can be concluded that Mendeley readership can indeed play a role as an alternative approach (to journal-based impact indicators) to find highly-cited outputs, being the only one of all altmetric sources exhibiting such possibility.

Although we haven't approached the issue of prediction of later highly cited publications, as it would be necessary to study early readership counts (which are currently not available in the data provided by Mendeley); it could be argued that this good filtering ability of Mendeley readership could be seen also as a strong indication of a potential predictability of future highly cited publications, particularly if we take into account its faster uptake (i.e., Mendeley readership are accumulated earlier than citations). Therefore, as suggested by Thelwall & Sud (2015) "future work with early Mendeley reader counts and later citation counts for the same set of articles is urgently needed to check this hypothesis" of whether Mendeley readership can predict future citations, and in this case, also highly cited papers.

5.5. Final remarks

The results of this study show that Mendeley readership scores are an effective tool to filter highly cited publications. This result, together with the moderate correlations between citations and readership found in previous studies (Thelwall & Wilson, 2015; Haustein et al., 2014b) as well as the "pre-citation role"[11] that is expected from Mendeley readership (i.e., that Mendeley users save documents in their libraries to cite them later, cf. Haustein, Bowman & Costas, 2015; Thelwall & Sud, 2015) make it possible to argue that Mendeley readership and citations are two different but connected processes that could be capturing a similar type of impact. However, from a more conceptual point of view, saving a document in Mendeley and citing it are two fundamentally different acts (Haustein, Bowman & Costas, 2015). Thus, considering the broad spectrum of reasons why Mendeley users may save documents in their libraries (for example, not only to cite them later, but also to use them for reading, teaching, self-awareness, individual non-academic interests, personal curiosity, etc.), it would not be correct to fully assimilate Mendeley readership impact to citation impact. Mendeley users cannot be expected to adhere to the same norms and expectations when they save a document as when they cite it[12] and clearly more research is necessary in order to better understand the differences and similarities between these two metrics.

[11] Results of a survey on Mendeley showed that 85% of respondents have saved documents in Mendeley to cite them later (Mohammadi, Thelwall, & Kousha, 2015), which would support the idea of Mendeley readership as a "pre-citation" event (cf. Haustein et al., 2015).

[12] For instance, Mendeley users don't necessarily follow the Mertonian norms of "communism", "universalism", "disinterestedness" and "organized skepticism" (Merton, 1973) as pointed out by Haustein et al. (2015) when they select a document to be saved in their libraries, while they could be more driven by these norms when selecting a document for citation.

Finally, there are also important technical issues (e.g., differences between the bibliographic metadata reported by Mendeley and WoS) that need to be considered and that can influence the data retrieval and the matching of records based on different identifiers (such as DOI, titles, journals, publication years, etc.) and hence can have an influence in the number of readership per publication (Thelwall, 2015; Zahedi, Bowman, & Haustein, 2014).
Although this study emphasizes the ability of Mendeley readership to identify highly cited publications and its role as a potential evaluative tool, more research is necessary to explore the abovementioned issues and limitations as well as to reveal more accurately the meaning of Mendeley readership and its potential value for research evaluation purposes. Follow up research should continue to explore the conceptual meaning of Mendeley readership and its relationship with citation indicators, as well as study whether Mendeley readership can be used to predict future citation. The disciplinary differences in the database coverage on which the citation and readership data are based is an important factor that should be considered when interpreting the results, and further research should focus on determining the potential influence that different levels of coverage may have on the value of Mendeley readership over journal indicators for all disciplines of Science.

5.6. Acknowledgment

This paper is an extended version of a paper accepted for oral presentation at the 15th International Conference on Scientometrics and Informetrics (ISSI), 29 Jun-4 July, 2015, Bogazici University, Istanbul (Turkey). The authors are grateful to Henri de Winter (CWTS) for his support on the Mendeley data collection and data management for this study. Also, special thanks to Ludo Waltman, Alex Rushforth (CWTS) and the anonymous referees of the journal for their valuable comments on this paper. Zohreh Zahedi was partially funded by the Iranian Ministry of Science, Research, & Technology scholarship program (MSRT grant number 89100156).

5.7. References

Abramo, G., & D'Angelo, C. A. (2015). Ranking research institutions by the number of highly-cited articles per scientist. *Journal of Informetrics*, 9(4), 915–923. https://doi.org/10.1016/j.joi.2015.09.001

Adler, R., Ewing, J., & Taylor, P. (2008). Citation statistics. A report from the International Mathematical Union. Available from: www.mathunion.org/publications/report/citationstatistics0.

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159-170. https://doi.org/ 10.3152/147154403781776645

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, 23(2): 166-173.https://doi.org/ 10.1093/reseval/rvu002

Bornmann, L., & Daniel, H.D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H. D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, 8(1), 93-102.

Collins, E., Bulger, M. E., & Meyer, E. T. (2012). Discipline matters: Technology use in the humanities. *Arts and Humanities in Higher Education*, 11(1–2), 76–92.

Costas, R., Zahedi, Z., & Wouters, P. (2015a). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66: 2003–2019. https://doi.org/ 10.1002/asi.23309

Costas, R., Zahedi, Z., & Wouters, P. (2015b). The thematic orientation of publications mentioned on social media: Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, Vol. 67 Iss: 3, pp.260 - 288.

Crespo, J. A., Herranz, N., Li, Y., & Ruiz-Castillo, J. (2014), The effect on citation inequality of differences in citation practices at the web of science subject category level. *Journal of the Association for Information Science and Technology*, 65: 1244–1256. https://doi.org/ 10.1002/asi.23006

Crespo, J.A., Li, Y., & Ruiz–Castillo, J. (2013). Correction: The Measurement of the Effect on Citation Inequality of Differences in Citation Practices across Scientific Fields. *PLoS ONE* 8(5).https://doi.org/10.1371/annotation/d7b4f0c9-8195-45de-bee5-a83a266857fc

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, 35 (1): 31-44. https://doi.org/ 10.1016/S0306-4573(98)00028-4.

Gruzd, A., Staves, K., & Wilk, A. (2012). Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model. *Computers in Human Behavior*, 28(6), 2340–2350. https://doi.org/10.1016/j.chb.2012.07.004

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, 101(2), 1419-1430. https://doi.org/10.1007/s11192-014-1261-3

Harman, D. (2011). Information retrieval evaluation. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 3(2), 1-119.

Hauff, C. & Houben, G.J.(2011). Deriving Knowledge Profiles from Twitter. In M Wolpers, C Delgado Kloos & D Gillet (Eds.), *Sixth European Conference on Technology Enhanced Learning, EC-TEL 2011 Towards Ubiquitous Learning*, Vol. 6964. Lecture Notes in Computer Science (pp. 139-152). Berlin, Germany: Springer Verlag.

Haunschild, R., & Bornmann, L. (2016).Normalization of Mendeley reader counts for impact assessment. *Journal of Informetrics*, 10(1): 62-73. https://doi.org/10.1016/j.joi.2015.11.003

Haunschild, R., & Bornmann, L. (2015). For which disciplines are papers covered in F1000Prime interesting? An analysis of discipline-specific reader data from Mendeley. available in Figshare (submitted to F1000 research on 19/01/2015 2015).

Haunschild, R., Bornmann, L., & Leydesdorff, L. (2015). Networks of reader and country status: An analysis of Mendeley reader statistics. arXiv preprint arXiv:1504.07482.

Haustein, S., Bowman, T. D., & Costas, R. (2015). Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics: A Festschrift in Honor of Blaise Cronin* (pp. 1–24). De Gruyter Mouton.

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014 a). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656-669. https://doi.org/10.1002/asi.23101

Haustein, S., & Larivière, S. (2014). Mendeley as a Source of Readership by Students and Postdocs? Evaluating Article Usage by Academic Status. In *Proceedings of the the 35th International Association of Technological University Libraries (IATUL)*, 2-5 June 2014, Helsinki, Finland.

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters. (2014b). Tweets vs. Mendeley readers: How do these two social media metrics differ? *Information Technology*, 56(5):207–215. https://doi.org/10.1515/itit-2014-1048

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics,* 101(2), 1027-1042. https://doi.org/ 10.1007/s11192-014-1229-3

Ivanović, D., & Ho, Y. S. (2014). Highly cited articles in the Information Science and Library Science category in Social Science Citation Index: A bibliometric analysis. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/0961000614537514.

Jiménez-Contreras, E., de Moya Anegón, F., & López-Cózar, E. D. (2003). The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of

Research Activity (CNEAI). *Research policy*, 32(1), 123-142. https://doi.org/ 10.1016/S0048-7333(02)00008-2

Korevaar, J.C., & Moed, H.F. (1996). Validation of bibliometric indicators in the field of mathematics. *Scientometrics*, 37(1):117–30.

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S. & Curry, S. (2016). A simple proposal for the publication of journal citation distributions. *bioRxiv*, 062109. https://doi.org/10.1101/062109

Levitt, J. M., & Thelwall, M. (2011). A combined bibliometric indicator to predict article impact. *Information Processing & Management,* 47(2), 300-308.

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. https://doi.org/ 10.1002/asi.23369

Mas-Bleda, A., Thelwall, M., Kousha, K. and Aguillo, I.F. (2014) .Do Highly highly Cited researchers successfully use the social web? *Scientometrics*, 101 (1): 337-356.

Meho, L. I., & Sugimoto, C. R. (2009). Assessing the scholarly impact of information studies: A tale of two citation databases: Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 60(12), 2499-2508.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131-149.

Mohammadi, E., Thelwall, M., & Kousha, K. (2015). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23477

Nederhof, T. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, 66(1), 81–100.

Ortega, J. L. (2015). Disciplinary differences in the use of academic social networking sites. *Online Information Review*, 39(4): 520 - 536.

Rushforth, A., & Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands.*Minerva*, 53(2), 117-139.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43 (9): 628–638.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal(BMJ*), 314 :498-502.

Schneider, J.W., & Costas, R. (2014). Identifying potential 'breakthrough' research articles using refined citation analyses: Three explorative approaches. In *Proceedings of the 19th Science and Technology Indicators Conference*, 3-5 September, Leiden University, The Netherlands.

Stern, D. I. (2014). High-ranked social science journal articles can be identified from early citation information. PloS one, 9(11), e112520.

Thelwall, M. (2015). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? *Journal of Librarianship and Information Science*. https://doi.org/10.1177/0961000615594867

Thelwall, M. & Fairclough, R. (2015). Geometric journal impact factors correcting for individual highly cited articles. *Journal of Informetrics,* 9(2): 263-272. https://doi.org/10.1016/j.joi.2015.02.004.

Thelwall, M. & Kousha, K. (2014), Academia.edu: Social network or academic network? *Journal of the Association for Information Science and Technology*, 65 (4):721–731.

Thelwall, M., & Maflahi, N. (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. J*ournal of the Association for Information Science and Technolog*y, 66: 1124–1135. https://doi.org/10.1002/asi.23252

Thelwall, M., & Sud, P. (2015). Mendeley Readership Counts: An Investigation of Temporal and Disciplinary Differences. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23559

Thelwall, M., & Wilson, P. (2015). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23501

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841.

Tijssen, R., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54(3), 381–397. https://doi.org/10.1023/A:1016082432660

Van Leeuwen, T., Moed, H., Tijssen, R., Visser, M., & Van Raan, A. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51(1), 335-346. https://doi.org/10.1023/A:1010549719484

Vieira, E.S. & Gomes, J.A.N.F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, 4(1):1-13. https://doi.org/10.1016/j.joi.2009.06.002

Waltman, L. (2016, July 11). The importance of taking a clear position in the impact factor debate [blog post]. Retreived from: www.cwts.nl/blog?article=n-q2w2c4&title=the-importance-of-taking-a-clear-position-in-the-impact-factor-debate

Waltman, L., & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445. https://doi.org/10.1002/asi23040

Waltman, L., & Van Eck, N.J. (2013). Source normalized indicators of citation impact: an overview of different approaches and an empirical comparison. *Scientometrics*, 96(3), 699-716. https://doi.org/10.1007/s11192-012-0913-4

Waltman, L., Van Eck, N., & Wouters, P. (2013). Counting publications and citations: Is more always better?. *Journal of Informetrics*, 7(3):635-641. https://doi.org/10.1016/j.joi.2013.04.001

Zahedi, Z., Costas, R., & Wouters, P. (2014a). How well developed are altmetrics? A cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. *Scientometrics*, 101(2), 1491–1513. https://doi.org/10.1007/s11192-014-1264-0

Zahedi, Z., Costas, R., & Wouters, P. (2014b). Assessing the impact of the publications read by the different Mendeley users: Is there any different pattern among users? In *Proceedings of the 35th International Association of Technological University Libraries (IATUL)*, 2-5 June 2014, Helsinki, Finland.

Zahedi, Z., Bowman, T., & Haustein, S. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. In *Metrics 2014, ASIS&T Workshop on Informetric and Scientometric Research*, Seattle, US.

Zahedi, Z., Costas, R., & Wouters, P. (2015). Do Mendeley readership counts help to filter highly cited WoS publications better than average citation impact of journals (JCS)? In *Proceedings of the 15th International conference on scientometrics & informetrics*, 29Jun-4July 2015, Bogazici University, Istanbul, Turkey.

CHAPTER 6

Exploring topics of interest of Mendeley users using VOSviewer overlay visualizations[1]

[1] Zahedi, Z., and Van Eck, Nees (Accepted). Exploring topics of interest of Mendeley users using VOSviewer overlay visualizations. Journal of Altmetrics.

Abstract

This chapter presents a fine-grained overview of the usage behavior and topics of interest of different types of users in Mendeley. The analysis is based on 1.2 million Web of Science indexed publications published in 2012. The disciplinary differences in the reading (saving) patterns of different types of Mendeley users are identified and depicted using VOSviewer overlay visualizations. The findings show that compared to other fields, publications from Mathematics & Computer Science have the lowest coverage in Mendeley. In contrast, publications from the Social Sciences & Humanities receive on average the highest number of readers in Mendeley. Although the highest uptake of Mendeley is by students, this differs across fields. Professors, students, and librarians are mainly active in the Social Sciences & Humanities, a field of science with a relatively low citation density in Web of Science. In contrast, researchers and other professionals are mainly active in fields with a relatively high citation density such as the Biomedical & Health Sciences and the Life & Earth Sciences. In addition, it seems that researchers and professionals are relatively more interested in practical, methodological, and technical oriented topics while professors and students are attracted by the more educational and theoretical oriented topics. These different usage patterns among user types possibly reflect the way in which scholarly publications are used for scientific, educational, or other professional purposes. This information could inform relevant stakeholders, such as researchers, librarians, publishers, funders, and policy makers of the scientific, educational, or professional values of publications.

Keywords

Altmetrics; Fields of science; Mendeley; Publication-level classification; Readership statistics; Topics; user behaviour; Visualization; VOSviewer

6.1. Introduction and Background

The social reference manager tool Mendeley is a prevalent source of altmetric data. It is known that the coverage, density, and distribution of Mendeley readership[1] varies substantially across disciplines (Costas, Zahedi, & Wouters, 2015). Depending on the field, Mendeley covers 45% to 90% of the publications in the Scopus database (Thelwall & Sud, 2016), 60% to 90% of the publications in the Web of Science (WoS) database (Zahedi, Costas, & Wouters, 2017), and more than 80% of the publications published by PLOS (Priem, Piwowar, & Hemminger, 2012). Fields from the Social Sciences & Humanities (such as Sociology, Communication, Business, Psychology, Anthropology, Educational Research, and Linguistics) have a relatively high coverage and a relatively high number of readers in Mendeley (Costas, Zahedi, & Wouters, 2015; Hammarfelt, 2014; Mohammadi & Thelwall, 2014). In contrast, fields from Mathematics & Computer Science (such as Analysis, Algebra and Number Theory, Geometry and Topology) show a relatively low coverage and a relatively low number of readers (Thelwall, 2017; Zahedi, Costas, & Wouters, 2014a). Moreover, readership and citation counts per publication have similar skewed distributions across different fields of science (Costas, Haustein, Zahedi, & Larivière, 2016). Hence similar to citations, normalization approaches for correcting field differences for Mendeley readership have been suggested (Costas, Perianes-Rodríguez, & Ruiz-Castillo, 2017; Haunschild & Bornmann, 2016).

In some previous studies, the readership activity of Mendeley users has been analyzed based on the self-declared academic disciplines of users. For example, co-readership based on the publications in the libraries of users and the self-declared academic disciplines of users have been used to measure and depict the similarity of subject areas within the field of Educational Technology (Kraker, Schlögl, Jack, & Lindstaedt, 2015). The analysis of the network of co-readers in Mendeley also showed that students and postdocs in Mendeley have more common topical interests than other types of users in Mendeley (Haunschild, Bornmann, & Leydesdorff, 2015). In other studies, existing field classification systems have been used to compare readership between different types of users across different fields of science. The readership activity of Mendeley users has, for instance, been analyzed using the 5 main disciplines and 22 sub-disciplines from the NSF classification system (Haustein & Larivière, 2014; Mohammadi, Thelwall, Haustein, and Larivier, 2015), the 250 subject categories available in the WoS database (Zahedi & Van Eck, 2015), and the 310 subject areas available in the Scopus database (Thelwall, 2017). The results of these studies show that substantial differences in readership practices between (sub)fields and user types exist. Moreover, the extent to which the number of readers correlates with the number of citations varies across different (sub)fields and between user types.

Most of the previous studies are based on restricted Mendeley data (only top three user types per publication) and focus on broad fields of science. It is not known yet how readership per

[1] The act of saving a document by different users in Mendeley is commonly known as 'readership' and therefore the same terminology is used in this paper to refer to this activity. Even though it does not always mean that the user who has 'saved' the document in his/her own private library has already read the document or vice versa.

user type varies across detailed micro-level fields and how these user types differ in their topics of interest. This is the first large scale and systematic analysis of readership activity across detailed micro-level fields in which complete data on the readership activities of Mendeley users is taken into account. Also, in addition to the overall readership activities of Mendeley users, the relative activity of different types of Mendeley users has been considered in this study. In this way, we have been able to uncover the topics on which different types of users focus relatively strongly. Moreover, a new view on readership statistics has been introduced by looking at the number of readers of publications normalized by the number of citations received by those publications. Combining these different usage statistics and patterns among user types provides insight into the way in which scholarly publications are used for scientific, educational, training, and other practical purposes. In this way, readership statistics could be used by relevant stakeholders (researchers, librarians, publishers, funders, policy makers, etc.) to get more insight into the full impact of scholarly publications. Hence, to determine whether information from Mendeley would be helpful in this respect, we will address the following main research questions in this paper:

1. What is the overall and relative readership activity of different types of Mendeley users across research fields? In which research fields are the different types of Mendeley users relatively most and least active?
2. What are the topics of interest within research fields? Are there any differences between Mendeley user types? Could such differences reflect different types of usage of scholarly publications?

This paper is organized as follows. We first describe our dataset and analysis methods. Results are then reported. The paper concludes with a summary, a discussion of some key observations, and suggestions for additional work.

6.2. Data & Methodology

This study is based on a dataset of 1,196,226 publications collected from the WoS database. The dataset includes all publications of the document types 'article' and 'review' published in 2012 with a Digital Object Identifier (DOI).[2] The DOIs of the collected publications were used to retrieve readership data from Mendeley by using the Mendeley REST API in July 2016[3]. This readership data also includes information on the 'academic status' of users as indicated by the

[2] About 13% of the 2012 WoS indexed articles and reviews do not have a DOI and are therefore excluded from the analysis.
[3] It is suggested in the literature (Zahedi, Haustein, & Bowman, 2014) that the best strategy to retrieve readership data using the Mendeley API is to perform searches which are based on a combination of DOIs and article titles. In this paper, however, we have chosen to search and match publications based on DOI only in order to keep our data collection accurate and transparent. Matching based on DOI only may lead to missed matches but will avoid wrong matches.

users in their Mendeley profile[4]. To minimize the effect of national and disciplinary differences between the designations of academic and professional appointments and positions, users were grouped into five broad user types. Based on their 'academic status', users were grouped into the following user types:

- *Students*: students (Bachelor), students (Master), students (postgraduate), doctoral students, and PhD students.
- *Researchers*: postdocs, researchers (at non-academic institutions), and researchers (at academic institutions).
- *Professors*: assistant professors, associate professors, professor, lecturers, and senior lecturers.
- *Librarians*: librarians or other library professionals.
- *Other professionals*: other professionals (including medical doctors, nurses, nutritionists, lawyers, etc.).

For each publication, readership counts for all users and readership counts for individual user types were calculated. Citations were counted until the end of week 26 (July) of 2016 using the in-house version of the WoS database of the Centre for Science and Technology Studies of Leiden University. The publications in the dataset have been assigned to 4,113 micro-level fields and to five main fields of science. The 4,113 micro-level fields have been constructed algorithmically based on 282.4 million citation relations between 17.8 million publications from the period 2000–2015 indexed in the WoS database (Waltman & Van Eck, 2012). The definitions used in the 2016 version of the CWTS Leiden Ranking[5] have been used to aggregate the 4,113 micro-level fields into five main fields of science. Table 1 provides the number of publications in our dataset assigned to each of the five main fields of science.

Table 1. Description of the dataset.

	No. pub.
All fields	1,196,226
Biomedical & Health Sciences	471,809
Life & Earth Sciences	172,260
Mathematics & Computer Science	87,005
Physical Sciences & Engineering	355,163
Social Sciences & Humanities	91,953

Figure 1 shows a visualization that provides an overview of the 4,113 micro-level fields and the five main fields of science used in this study. Each circle represents a micro-level field. The size of a circle indicates the number of publications in our dataset in a micro-level field. The

[4] The 'academic status' is self-declared by Mendeley users. It therefore may happen that users forget to update or simply do not update their 'academic status' in their Mendeley profile when it has been changed, e.g., due to a job change or promotion. This should be kept in mind while interpreting the results of this study.
[5] See http://www.leidenranking.com/information/fields for more information.

larger the circle, the larger the number of publications in our dataset. The distance between two circles approximately indicates the relatedness of two micro-level fields, where the relatedness is determined by citation relations between the fields. In general, the smaller the distance between two circles, the stronger the micro-level fields are related to each other. The color of a circle indicates the main field to which a micro-level field belong. The color coding and positioning of the main fields is as follows. Mathematics & Computer Science (purple) are located in the top-right, the Physical Sciences & Engineering (blue) are located in the bottom-right, the Life & Earth Sciences (yellow) are located in the center, the Biomedical & Health Sciences (green) are located in the bottom-left, and the Social Sciences & Humanities (red) are located in the top-left in the visualization.
The 4,113 micro-level fields and the five main fields of science enabled us to analyze readership activity in Mendeley at different levels of granularity. In order to analyze readership activity from different perspectives and to allow for analyzing differences between research fields and user types, the following statistics have been calculated for each of the micro-level fields and main fields of science:

- *Publications*: Number of publications.
- *Mendeley coverage (all users / user type)*: Total number and percentage of publications saved by all Mendeley users or a specific Mendeley user type.
- *Citations (all users / user type)*: Total number and average number of citations of publications saved by all Mendeley users or a specific Mendeley user type.
- *Readers (all users / user type)*: Total number and average number of readers per publication based on the activity of all Mendeley users or a specific Mendeley user type.
- *Normalized readers (all users / user type)*: Share of the number of readers based on the activity of all Mendeley users or a specific Mendeley user type in a field divided by the share of the number of citations in a field.
- *Relative activity (user type)*: Average number of readers per publication based on the activity of a specific Mendeley user type divided by the average number of readers per publication based on the activity of all Mendeley users.

Visualizations providing overviews of the above described statistic at the level of 4,113 micro-level fields were constructed. The VOSviewer software tool (Van Eck & Waltman, 2010) was used for this purpose. So-called overlay visualizations were constructed using version 1.6.6 of the VOSviewer software tool. These visualizations can be used to show additional information on top of a base map (e.g., Leydesdorff & Rafols, 2012; Van Eck, Waltman, Van Raan, Klautz, & Peul, 2013). In this case, the visualization of the 4,113 micro-level fields presented in Figure 1 was used as a base map. The constructed overlay visualizations enabled us to analyze readership activity in Mendeley in a fine-grained way and to identify possible differences between fields and user types.

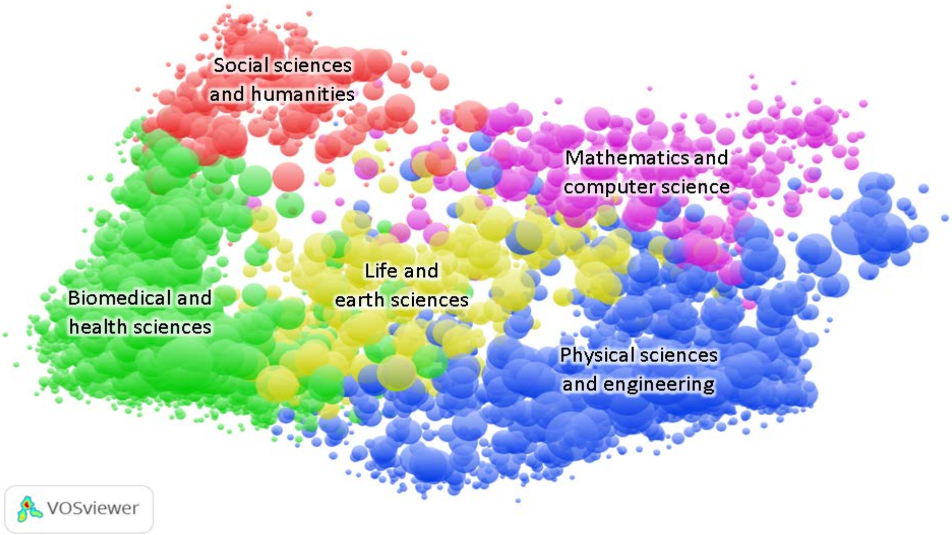


Figure 1. Visualization of the CWTS publication-level classification system used in this study. Each circle represents one of the 4,113 micro-level fields. The size of a circle indicates the number of publications in our dataset in a micro-level field. The larger the circle, the larger the number of publications. The color of a circle indicates the main field to which a micro-level field belongs. The color coding is as follows: Biomedical & Health Sciences (green), Life & Earth Sciences (yellow), Mathematics & Computer Science (purple), Physical Sciences & Engineering (blue), and Social Sciences & Humanities (red).

6.3. Results

A number of different analyses were performed in order to answer the research questions stated in the introduction of this paper. This section presents the results of these analyses. First, results on the coverage of publications by field and user type are presented in order to provide a complete overview of the coverage of Mendeley. Then, results on the readership activity in Mendeley by field and user type are presented. Finally, results on the topics of interest of Mendeley users are presented.

As already indicated in the Data and Methodology section, overlay visualizations of the 4,113 micro-level fields played an important role in our analyses. In this section, static figures of the overlay visualizations are presented. The overlay visualizations can also be explored interactively using the VOSviewer software tool[6]. The interactive version of the overlay visualizations is available online at https://goo.gl/CJVRzL. The interactive visualizations offer the possibility to zoom in on a specific area in the visualizations and to explore in more detail the micro-level fields located in that area. The interactive visualizations also offer additional information that is not visible in the static figures. By hovering the mouse over a micro-level field, more detailed information on the field is presented.

[6] Please note that Java needs to be installed on your computer in order to access the interactive overlay visualizations in VOSviewer.

6.3.1. Coverage of publications saved by Mendeley users across different fields

In this subsection, we analyze the coverage of our dataset in Mendeley. Table 2 presents the total number of publications in our dataset and the coverage of these publications in Mendeley. With coverage in Mendeley we mean the percentage of publications with at least one reader in Mendeley. Table 2 also presents the breakdown of the coverage by field and by user type. In Figure 2, the coverage of the 4,113 micro-level fields is visualized using VOSviewer overlay visualizations. The coverage is shown for all Mendeley users (Figure 2a) and individual user types (Figure 2b, 2c, 2d, 2e, and 2f). As explained above, each circle represents a micro-level field. The size of a circle indicates the number of publications in our dataset in a micro-level field. The larger the circle, the larger the number of publications. The color of a circle indicates the percentage of publications in a micro-level field that is covered in Mendeley. The color ranges from blue to green to red showing low, medium, and high coverage. The positions of the main fields are as explained in the previous section.

As can be seen in Table 2 and Figure 2, publications from almost all fields are well covered in Mendeley. Publications from the Biomedical & Health Sciences (96.3%) and the Life & Earth Sciences (95.7%) have the highest coverage in Mendeley followed by those from the Social Sciences & Humanities (95.9%) and the Physical Sciences & Engineering (90.2%). Publications from Mathematics & Computer Science have the lowest coverage (79.9%). This is also clearly visible in Figure 2a by looking at the top-right side of the visualization.

Table 2. Coverage of WoS publications in Mendeley.

	All users		Professors		Researchers		Students		Librarians		Other professionals	
	Total	**Covg.**	**Total**	**Covg.**	**Total**	**Covg.**	**Total**	**Covg.**	**Total**	**Covg.**	**Total**	**Covg.**
All fields	1,108,574	92.7%	761,193	63.6%	840,955	70.3%	1,051,314	87.9%	119,276	10.0%	397,278	33.2%
Biomedical & Health Sciences	454,525	96.3%	323840	68.6%	371,405	78.7%	436,299	92.5%	63,908	13.5%	221,046	46.9%
Life & Earth Sciences	164,796	95.7%	123057	71.4%	140,964	81.8%	159,581	92.6%	18,812	10.9%	70,622	41.0%
Mathematics & Computer Science	69,524	79.9%	41773	48.0%	35,239	40.5%	61,389	70.6%	3,954	4.5%	11,227	12.9%
Physical Sciences & Engineering	320,342	90.2%	195276	55.0%	223,473	62.9%	299,027	84.2%	15,021	4.2%	59,974	16.9%
Social Sciences & Humanities	88,163	95.9%	72322	78.7%	65,566	71.3%	85,882	93.4%	16,769	18.2%	32,756	35.6%

Table 3. Reader statistics (total number and average number of readers per publication in Mendeley).

	All users		Professors		Researchers		Students		Librarians		Other professionals	
	Total	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**
All fields	21,018,929	19.0	2,564,289	3.4	4,228,402	5.0	13,253,718	12.6	155,441	1.3	813,505	2.0
Biomedical & Health Sciences	9,324,273	20.5	1,124,167	3.5	2,048,169	5.5	5,569,493	12.8	84,078	1.3	496,042	2.2
Life & Earth Sciences	4,033,234	24.5	447,990	3.6	920,281	6.5	2,489,761	15.6	22,858	1.2	151,851	2.2
Mathematics & Computer Science	755,850	10.9	101,849	2.4	96,154	2.7	536,024	8.7	4,701	1.2	17,046	1.5
Physical Sciences & Engineering	4,366,395	13.6	531,560	2.7	864,154	3.9	2,868,858	9.6	16,912	1.1	84,526	1.4
Social Sciences & Humanities	2,482,076	28.2	349,830	4.8	290,637	4.4	1,753,882	20.4	25,809	1.5	61,632	1.9

Table 4. Citation statistics (total number and average number of citations per publication in WoS).

	All users		Professors		Researchers		Students		Librarians		Professionals	
	Total	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**	**Total**	**Avg.**
All fields	12,689,428	11.4	10,645,127	14.0	11,367,502	13.5	12,474,552	11.9	1,907,387	16.0	6,739,679	17.0
Biomedical & Health Sciences	5,822,784	12.8	5,000,332	15.4	5,390,883	14.5	5,744,959	13.2	1,079,625	16.9	3,905,394	17.7
Life & Earth Sciences	1,761,887	10.7	1,544,610	12.6	1,659,639	11.8	1,745,320	10.9	291,260	15.5	1,042,447	14.8
Mathematics & Computer Science	400,029	5.8	293,746	7.0	261,607	7.4	376,936	6.1	30,541	7.7	102,194	9.1
Physical Sciences & Engineering	4,140,132	12.9	3,280,621	16.8	3,545,982	15.9	4,046,573	13.5	350,178	23.3	1,355,217	22.6
Social Sciences & Humanities	557,995	6.3	520,731	7.2	504,278	7.7	554,595	6.5	155,091	9.2	330,799	10.1

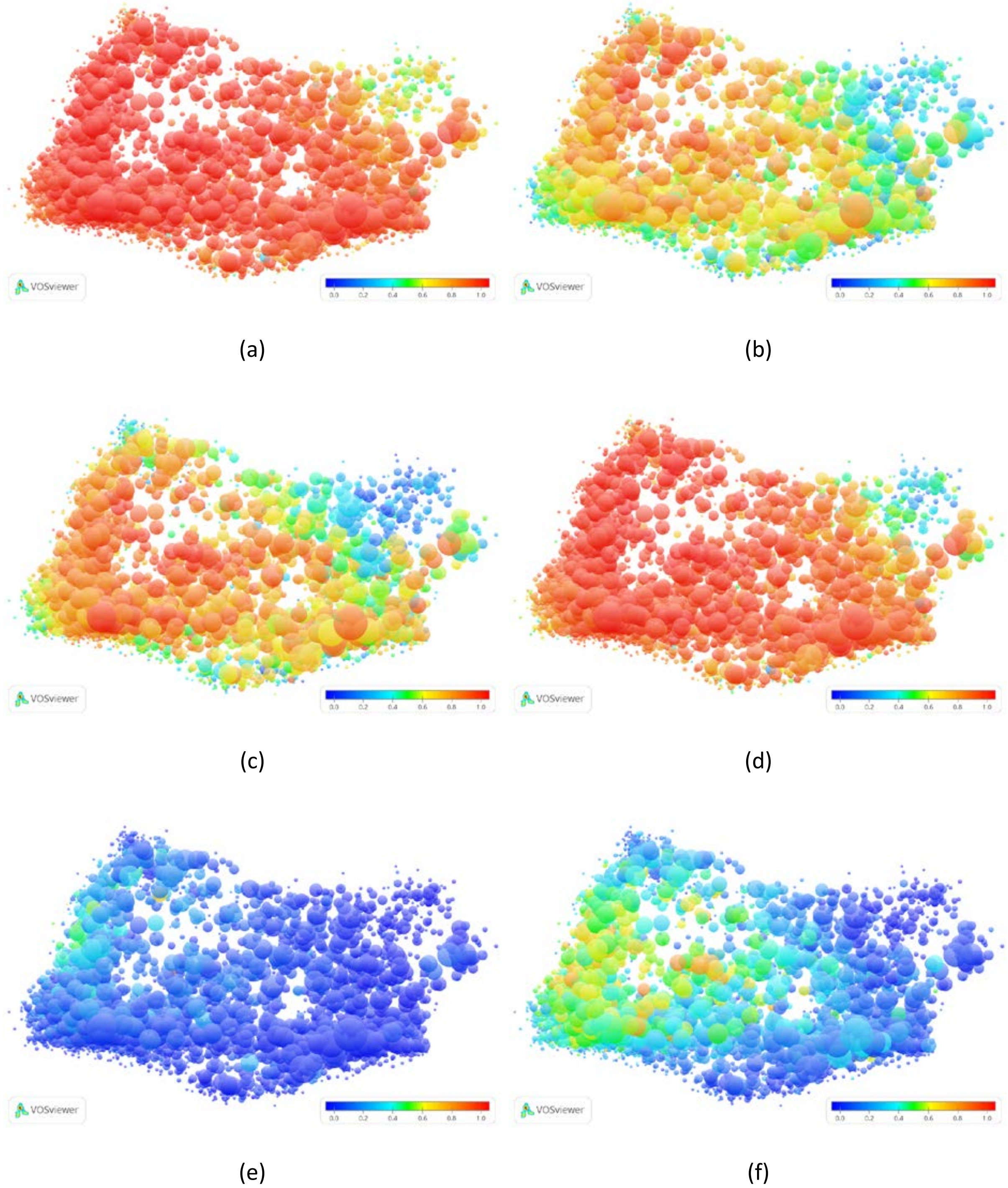


Figure 2. VOSviewer overlay visualizations showing the coverage of WoS publications in Mendeley per micro-level field. The coverage is shown for (a) all Mendeley users, (b) professors, (c) researchers, (d) students, (e) librarians, and (f) other professionals. An interactive version of the overlay visualizations is available online at https://goo.gl/CJVRzL.

Considering user types, it is clear that students are very active in Mendeley. Table 2 shows that the coverage is the highest for students (87.9%). This is followed by researchers (70.3%) and professors (63.6%). The lowest coverage is for other professionals (33.2%) and librarians (10.0%). The order of user types regarding the coverage is identical in the Biomedical & Health Sciences, the Life & Earth Sciences, and the Physical Sciences & Engineering. In Mathematics & Computer Science and the Social Sciences & Humanities, slightly more publications are saved by Mendeley users classified as professors than those identified as researchers. If we look at the coverage at the more detailed level of micro-level fields, we see that the visualization for students (Figure 2d) best resembles the general pattern based on all users (Figure 2a). The visualizations for professors (Figure 2b) and researchers (Figure 2c) are fairly comparable. A lower coverage can be observed for the peripheral micro-level fields and the micro-level fields from Mathematics & Computer Science. The visualizations for librarians (Figure 2e) and other professionals (Figure 2f) are most different from the general pattern (Figure 2a). Most micro-level fields show a relative low coverage for those user types. In the case of librarians, we see a somewhat higher coverage for micro-level fields at the intersection of the Biomedical & Health Sciences and the Social Sciences & Humanities. In the case of other professionals, micro-levels fields with the highest coverage are from the Biomedical & Health Sciences and the Life & Earth Sciences.

6.3.2. Reader and citation counts of publications

In this subsection, we analyze the reader counts and the citation counts of the publications in our dataset and we make comparisons across fields and between Mendeley user types. Table 3 presents the total and average number of readers per publication in Mendeley by main field and by Mendeley user type. Similarly, Table 4 presents the total and average number of citations per publication in WoS by main field and by Mendeley user type. Table 5 presents the normalized readership activity by main field and by Mendeley user type. Here, the number of readers in Mendeley are normalized by the number of citations in WoS. In Figure 3, the same statistics are presented at the level of the 4,113 micro-level fields using VOSviewer overlay visualizations. Each circle represents again a micro-level field and the size of a circle indicates the number of publications in our dataset in the corresponding micro-level field. The color of a circle indicates the average number of readers per publication (Figure 3a), the average number of citations per publication (Figure 3b), and the normalized number of readers in a micro-level field (Figure 3c).

Based on the results in Tables 3 and 4, differences between the number of readers and the number of citations across fields can be detected. It can be seen in Table 3 that publications from the Social Sciences & Humanities have on average the highest number of readers followed by publications from the Life & Earth Sciences, the Biomedical & Health Sciences, and the Physical Sciences & Engineering. In contrast, it can be seen in Table 4 that publications from the Physical Sciences & Engineering and the Biomedical & Health Sciences have on average the highest number of citations, followed by publications from the Life & Earth Sciences and the Social Sciences & Humanities. Publications from Mathematics & Computer

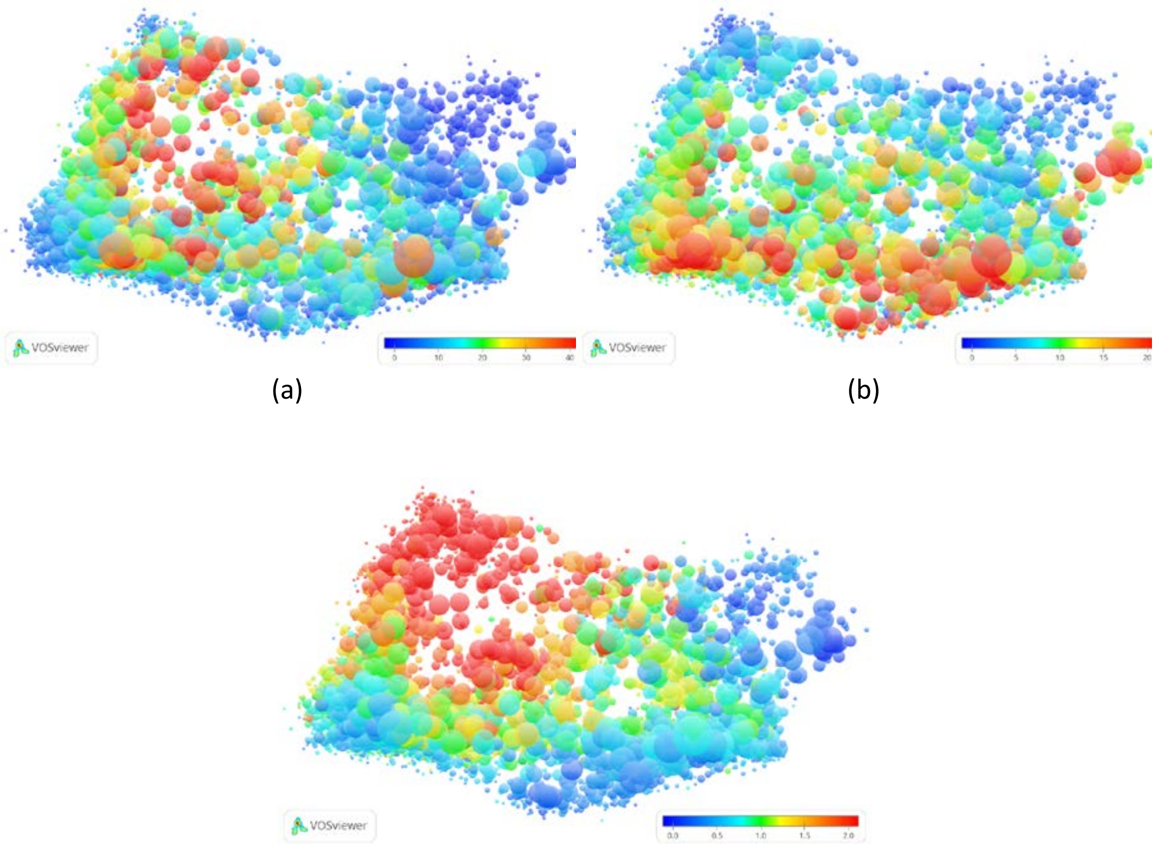
Science have on average both the lowest number of readers and the lowest number of citations. Nearly the same ordering of main fields can be observed for individual Mendeley user types. Researchers, librarians, and other professionals show some exceptions. In the case of researchers and other professionals, publications from the Life & Earth Sciences and the Biomedical & Health Sciences have on average a higher number of readers compared to publications from the Social Sciences & Humanities. In the case of librarians, publications from the Biomedical & Health Sciences have on average a slightly higher number of readers than publications from the Life & Earth Sciences.

If we look at the at the more detailed level of micro-level fields, we see that the visualization based on reader counts (Figure 3a) differs significantly from the visualization based on citation counts (Figure 3b). The differences that are visible in these visualizations are in line with the results at the level of main fields discussed previously. The highest average number of readers can be observed for micro-level fields from the Social Sciences & Humanities and the Life & Earth Sciences, while the highest average number of citations can be observed for micro-level fields from the Biomedical & Health Sciences and the Physical Sciences & Engineering.

Table 5 and Figure 3c show that particularly when the citation density of each main field and each micro-level field is considered, an above average number of readers for publications from the Social Sciences & Humanities is observable. This emphasizes the fact that fields with a relatively low citation density in WoS receive a relatively high number of readers in Mendeley. This is in particular observable for publications saved by librarians, professors, and students and could reflect the usefulness of these publications in practical, training, or educational contexts.

Table 5. Normalized reader statistics (number of readers in Mendeley normalized by the number of citations in WoS).

	All users	**Professors**	**Researchers**	**Students**	**Librarians**	**Other professionals**
All fields	1.7	0.2	0.4	1.1	0.1	0.1
Biomedical & Health Sciences	1.0	1.0	1.1	0.9	1.2	1.3
Life & Earth Sciences	1.4	1.3	1.6	1.4	1.1	1.4
Mathematics & Computer Science	1.1	1.2	0.7	1.2	0.9	0.6
Physical Sciences & Engineering	0.6	0.6	0.6	0.7	0.3	0.3
Social Sciences & Humanities	**2.7**	**3.1**	1.6	**3.0**	**3.8**	1.7



(a) (b)

(c)

Figure 3. VOSviewer overlay visualizations showing per micro-level field (a) the average number of readers in Mendeley, (b) the average number of citations in WoS, and (c) the number of readers in Mendeley normalized by the number of citation in WoS. An interactive version of the overlay visualizations is available online at https://goo.gl/CJVRzL.

6.3.3. Relative activity of Mendeley user types

In this subsection, the relative activity of Mendeley users is presented. Figure 4 provides visualizations of the relative activity of the different Mendeley user types at the level of the micro-level fields. As explained in the Data and Methodology section, the relative activity of a Mendeley user type in a field is calculated as the average number of readers per publication based on the activity of the Mendeley user type in the field divided by the average number of readers per publication based on the activity of all Mendeley users in the field. This approach provides us with a detailed overview of the micro-level fields in which different user types are relatively most and least active.

Based on the results in Figure 4 and in answer to the first main research question, it seems that professors (Figure 4a) are relatively seen the most active users in micro-level fields from Mathematics & Computer Science and some micro-level fields from the Social Sciences & Humanities and the Biomedical & Health Sciences. In contrast to professors, researchers (Figure 4b) are relatively more focused on micro-level fields from the Physical Sciences & Engineering, the Life & Earth Sciences, and the Biomedical & Health Sciences. Students (Figure

4c) mainly focus on the micro-level fields from the Social Sciences & Humanities, Mathematics & Computer Science, and the Physical Sciences & Engineering. Librarians (Figure 4d) are relatively most active in micro-level fields from the Biomedical & Health Sciences and the Social Sciences & Humanities, but their activity is also scattered across other micro-level fields. Other professionals (Figure 4e) are relatively most focused on micro-level fields from the Biomedical & Health Sciences and the Life & Earth Sciences.

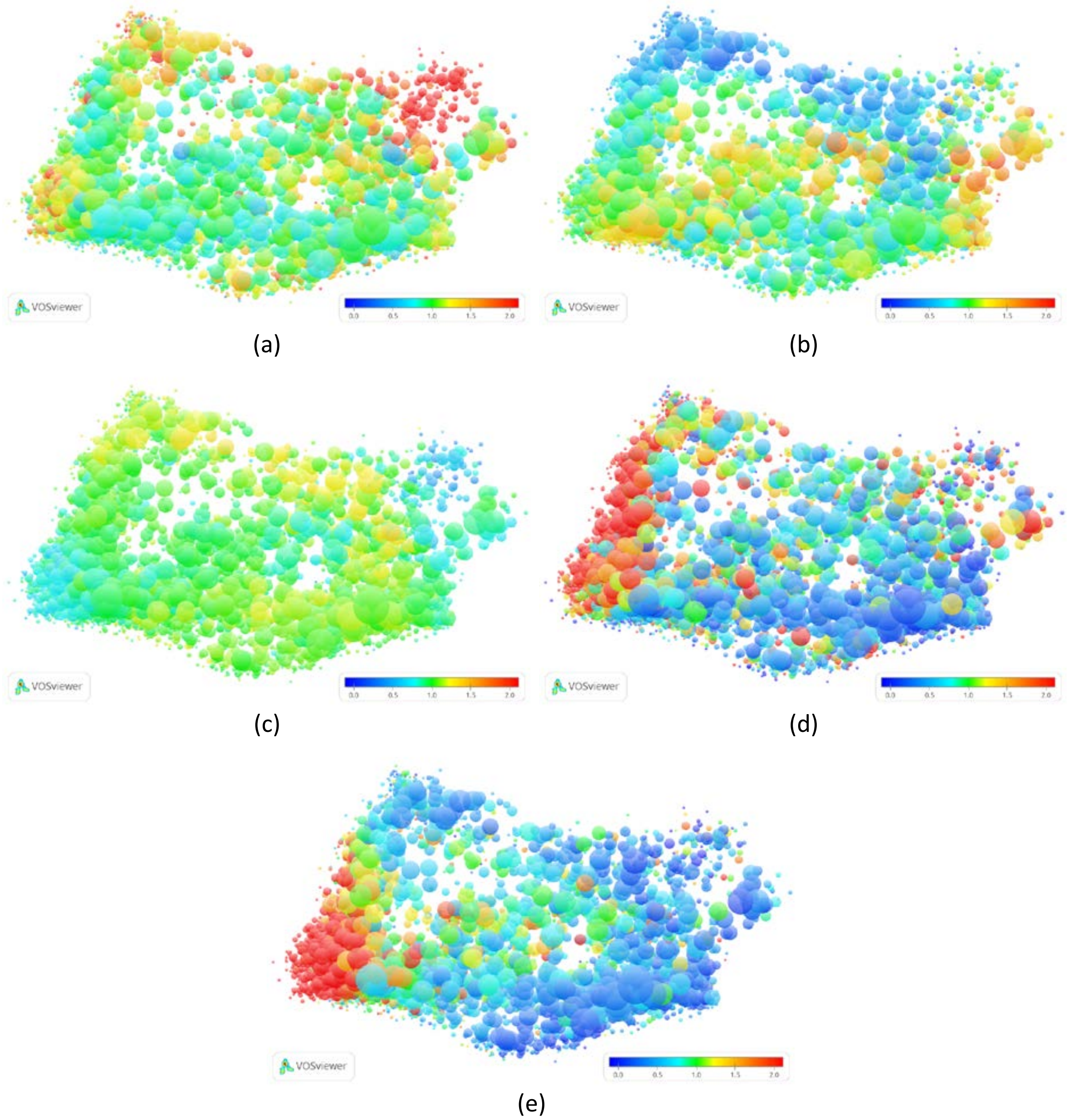


Figure 4. VOSviewer overlay visualizations showing the relative activity of Mendeley users across all micro-level fields. The relative activity is shown for (a) professors, (b) researchers, (c) students, (d) librarians, and (e) other professionals. An interactive version of the overlay visualizations is available online at https://goo.gl/CJVRzL.

6.3.4.Topics of interest of Mendeley user types

In this subsection, we present an overview of specific topics of interest of different Mendeley user types. For each user type, we identified the micro-level fields in which the users are relatively seen most active. The identification of the micro-level fields was done based on the relative activity of the users of a user type. Micro-level fields with a small absolute number of readers have been filtered out. To get an impression of the topics of the micro-level fields that have been selected in this way, a summary of the top 5 micro-level fields per Mendeley user type is provided in Tables A1 to A5 in the Appendix. For each micro-level field, the tables list the number of readers based on the activity of the corresponding user type, five characteristic terms, the three journals with the largest number of publications, and the most frequently saved publication.

By analyzing the micro-level fields listed in Tables A1 to A5 in the Appendix, differences between topics of interests of different Mendeley user types can be observed. Based on the results and in answer to the second main research question, it is interesting to see that professors have a relatively strong focus on topics related to teaching and education, like higher education, medical education, and second language acquisition (Table A1). Researchers seem to be interested in a broad range of topics. Their topics of interest range from climate research, pharmaceutical research, and biotechnology to astronomy and astrophysics (Table A2). Students seem to be biased towards topics such as business, management, and leadership (Table A3). Librarians show relatively most interest in topics that seem to be directly related to their work, namely bibliometrics and scientometrics, library science, and research utilization (Table A4). Other professionals seem to be mostly focused on biological, medical, and clinical oriented topics (Table A5). Below we further elaborate on the way in which different topic interests of different Mendeley user types could indicate different types of usage of scholarly publications.

6.4. Discussion and Conclusions

Mendeley is known as a promising source for altmetrics. Readership data from Mendeley can be used to reveal differences in reading (saving) behavior of different types of users. In this study, we have explored the usage of 1.2 million WoS indexed publications by different user types in Mendeley. The aim was to see if there are any differences in readership activity and topics of interest. VOSviewer overlay visualizations have been used to identify and depict these differences.

The findings of this study show that there are quite some disciplinary differences in terms of readership activity and in terms of the topics of interest among different user types in Mendeley. Publications from the Social Sciences & Humanities receive on average the highest number of readers, which may indicate that Mendeley is relatively more popular in this field than in other fields. It is interesting to see that this is in sharp contrast to citations, which are typically less concentrated in the Social Sciences & Humanities and most concentrated in the Biomedical & Health Sciences, the Life & Earth Sciences, and the Physical Sciences &

Engineering. The purpose for which Mendeley is used by different user types could help to explain the disciplinary differences. A recent study, for instance, found that professional health-related areas receive high readership in Mendeley due to the usage of the tool in training, while mathematics and high energy physics receive low readership due to use of other tools such as LaTeX (Thelwall, 2017). Another study showed that F1000 publications with the tag 'good for teaching' (publications that provide a good overview of a particular topic) receive most attention by Mendeley users classified as lecturers and publications with the tag 'new findings' receive most attention by users classified as researcher (Bornmann & Haunschild, 2015). It indicates that Mendeley is used for different purposes. The results of a survey among Mendeley users show that most of the respondents use Mendeley as a tool to cite literature in their publications, to keep track of relevant publications for their jobs, and to teach (Mohammadi et al.., 2014). Another recent survey shows that browsing papers and groups and connecting with other users are among the motivations of Mendeley users. Most of these users reported that they have read and cited or intended to cite most of the items in their Mendeley library (Chen, et al., 2018). Moreover, research-based features (managing documents and citations) in Mendeley are more popular by members of online groups than social-based features (making friends and connections) (Jeng , He & Jiang, 2015). These results are in line with the result of a survey that shows that the main reason for reading scientific publications in general for US faculty members is research and writing, teaching, current awareness, and education (Tenopir, King, Christian, & Volentine, 2015).

In terms of the topics of interest, the results of this study indeed indicate that different user types have relatively more attention for publications related to their role and the purpose for which they use Mendeley. This could range from conducting (literature) research, writing articles, training or (self) education, or the usage of a device or method. We have, for example, found that publications related to teaching and education show high readership among professors. This may be expected since professors use Mendeley among other things to organize literature for teaching and publishing. It assumes that professors reflect both educational and research use. Furthermore, the results seem to suggest that students focus on more general and fundamental topics. Although the fact that a publication is frequently saved by students does not provide conclusive evidence of its educational impact (Thelwall, 2016), students seem to frequently save fundamental and basic methodological publications that they read for educational purposes and this could reflect the use of Mendeley by students as a source for course material or as a source for their master or doctoral thesis. The strong interest of researchers in publications about applied sciences could show the scientific impact of these publications in an applied context. Researchers seem to focus on the research front and seem to use Mendeley mostly in a pre-citation context. Other professionals which may, for example, include medical doctors, nutritionists, and lawyers seem to be mostly interested in publications about diagnosis, treating, tools, and devices. In other words, other professionals are more likely to show interest in publications that have practical relevance in their work. It is also interesting to see that publications related to library and information sciences or clinical guidelines show high readership among librarians.

In conclusion, the various patterns of usage of scientific publications observed among the different user types in Mendeley could be an indication of the importance of these publications in research, training, (self) education, or in any professional, practical or applied context. Exploring how the topical interests of users differ across various fields provides useful information on who use scientific outputs, from which fields, and for what purposes. The possibility offered by Mendeley to track the use of scholarly publications by different types of users is an advantage that citation databases lack. Readership statistics based on different user types provide a broad overview of the usage of scholarly publications by a wide range of audiences including non-publishing users. This is important information in addition to information on citations, especially in fields with a low citation density or fields in which citations accumulate slowly. Detailed information on the usage of scholarly publications could help relevant stakeholders, such as researchers, librarians, publishers, funders, and policy makers to get more insight into the full impact of publications.

In addition to the number of users that have saved a publication, more information on specific activities of users in Mendeley could help to get a more accurate and comprehensive picture of the actual usage and impact of scholarly publications. For instance, it would be interesting to have more information on user actions such as assigned tags, notes that are added or parts that are highlighted in the full-text, and time spent on a saved publication. This type of information is not disclosed at the moment for Mendeley. More research is needed to find out whether and how this type of information can be useful in any practical application.

6.5. References

Bornmann, L., & Haunschild, R. (2015). Which people use which scientific papers? An evaluation of data from F1000: Discovery Service para Universidad de La Rioja. *Journal of Informetrics*, *9*(3), 477–487. https://doi.org/10.1016/j.joi.2015.04.001

Chen, P.-Y., Hayes, E., Larivière, V., & Sugimoto, C. R. (2018). Social reference managers and their users: A survey of demographics and ideologies. PLOS ONE, 13(7), e0198033. https://doi.org/10.1371/journal.pone.0198033

Costas, R., Haustein, S., Zahedi, Z., & Larivière. (2016). Exploring paths for the normalization of altmetrics: applying the Characteristic Scores and Scales. In *altmetrics16. Moving beyond counts: integrating context* (pp. 1–8), Romania, Bucharest.

Costas, R., Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2017). On the quest for currencies of science: field " exchange rates " for citations and Mendeley readership. *Aslib Journal of Information Management*. https://doi.org/10.1108/ AJIM-01-2017-0023

Costas, R., Zahedi, Z., & Wouters, P. (2015a). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. https://doi.org/10.1002/asi.23309

Costas, R., Zahedi, Z., & Wouters, P. (2015b). The thematic orientation of publications mentioned on social media Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, *67*(3), 260–288.
https://doi.org/10.1108/AJIM-12-2014-0173

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, *101*(2), 1419–1430. https://doi.org/10.1007/s11192-014-1261-3

Haunschild, R., & Bornmann, L. (2016). Normalization of Mendeley reader counts for impact assessment. *Journal of Informetrics*, *10*(1), 62–73.
https://doi.org/10.1016/j.joi.2015.11.003

Haunschild, R., Bornmann, L., & Leydesdorff, L. (2015). Networks of reader and country status: an analysis of Mendeley reader statistics. *PeerJ Computer Science*, *1*, e32.
https://doi.org/10.7717/peerj-cs.32

Haustein, S., & Larivière, V. (2014). Mendeley as a source of global readership by students and postdocs? Evaluating article usage by academic status. In *In 35th International Association of Technological University Libraries (IATUL), 2-5 June 2014, Aalto University, Helsinki, Finland. Helsinki* (pp. 1–10). Helsinki, Finland.

Jeng, W., He, D. & Jiang, J.(2015). User participation in an academic social networking service: A survey of open group users on Mendeley. *Journal of the Association for Information Science and Technology*, 66 (5): 890-904.

Kraker, P., Schlögl, C., Jack, K., & Lindstaedt, S. (2015). Visualization of co-readership patterns from an online reference management system. *Journal of Informetrics*, *9*(1), 169–182. https://doi.org/10.1016/j.joi.2014.12.003

Leydesdorff, L., & Rafols, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Informetrics*, *6*(2), 318–332.

https://doi.org/10.1016/j.joi.2011.11.003
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.
https://doi.org/10.1002/asi.23286
Mohammadi, E., Thelwall, M., Kousha, K., Cybermetrics, S., Science, C., & Street, W. (2014). Can Mendeley Bookmarks Reflect Readership ? A Survey of User Literature review Changes in scholarly reading habits in the digital era. *Journal of the Association for Information Science and Technology*, *67(5).* https://doi.org/10.1002/asi.23477
Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved from http://arxiv.org/abs/1203.4745
Tenopir, C., King, D. W., Christian, L., & Volentine, R. (2015). Scholarly article seeking, reading, and use: A continuing evolution from print to electronic in the sciences and social sciences. *Learned Publishing*, *28*(2), 93–105. https://doi.org/10.1087/20150203
Thelwall, M. (2016). Does Mendeley provide evidence of the educational value of journal articles? *Learned Publishing*, (August), 1–7. https://doi.org/10.1002/leap.1076
Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, *113*(3), 1721–1731. https://doi.org/10.1007/s11192-017-2557-x
Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, *8*(5), e64841.
https://doi.org/10.1371/journal.pone.0064841
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. https://doi.org/10.1002/asi.23559
Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
Van Eck, N. J., Waltman, L., Van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation Analysis May Severely Underestimate the Impact of Clinical Research as Compared to Basic Research. *PLoS ONE*, *8*(4), e62395. https://doi.org/10.1371/journal.pone.0062395
Waltman, L., & Van Eck, N. J. (2012). A New Methodology for Constructing a Publication-Level Classification System of Science, *Journal of the Association for Information Science and Technology, 63*(12), 2378–2392.
Waltman, L., & Van Eck, N. J. (2016). The Need for Contextualized Scientometric Analysis: An Opinion Paper. In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *STI Conference 2016. Peripheries, Frontiers and beyond* (pp. 1–9). Valencia, Spain.
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. https://doi.org/10.1007/s11192-014-1264-0
Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, *68*(10), 2511–2521. https://doi.org/10.1002/asi.23883
Zahedi, Z., & Van Eck, N. (2015). Identifying topics of interest of Mendeley users using the text mining and overlay visualization functionality of VOS viewer. In *20th International Conference in Science & Technology Indicators, 2-4, September, 2015, Lugano, Switzerland.*
Zahedi, Z., Haustein, S. & Bowman, T (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. In *SIGMET Metrics 2014 workshop*, *ASIS&T 2014 annual meeting*, 5 November 2014, Seattle, US.

CHAPTER 7

Discussion and conclusions

7.1. Introduction

Social media platforms have become increasingly popular among the general public as well as scholars. These platforms facilitate communication and distribution of scholarly publications across various audiences and hence play an important role in scholarly communications. Social media platforms allow (and record) different forms of events and interactions (e.g., tweets, retweets, likes, shares, commenting, savings, etc.) among their users and with all sorts of electronic objects (URLs, images, videos, etc.). According to Haustein, Bowman & Costas, (2016), the interactions of social media users with *research objects* are particularly relevant (including both *documents* – i.e., publications or research outputs; and *scholarly agents* - i.e., researchers, universities, funders, etc.). The user-generated data of the interactions of users with research objects and with other users in social media platforms have opened the possibility to develop new forms of indicators, in what has been known as *altmetrics*, and more specifically as *social media metrics* (Haustein, Bowman, & Costas, 2016; Wouters, Zahedi, & Costas, 2018). Social media metrics promised to provide diverse, fast, open, and comprehensive indicators of scientific activities (Priem, Taraborelli, Groth, & Neylon, 2010; Wouters & Costas, 2012), which could inform evaluators about relevant aspects of scientific performance. These new indicators have also gained significant attention in recent years, particularly in some national research assessment exercises (e.g., Research Excellence Framework (REF)[1] in the UK or the Dutch Standard Evaluation Protocol (SEP)[2]), since they were expected to reflect a broader perspective on the impact of research. However, their actual usefulness and possibilities for research evaluation are still unclear. The main research question that this PhD thesis addresses is *what is the potential usefulness and added value of social media metrics for informing research evaluation?*. The research presented in this PhD thesis has provided both empirical and conceptual answers to this main question.

In this chapter, a summary and discussion of the main findings of this thesis are presented. Specific answers are provided to the questions introduced in chapter 1. Conclusions based on the main findings of this thesis are also presented. Finally, reflections on future research in social media metrics are presented.

7.2. Main findings

Q1: What aspect of research impact do social media metrics reflect? In particular, how do social media metrics compare to more traditional bibliometric indicators?

Chapter 2 answers these questions through the analysis of the coverage (i.e., the percentage of publications with at least one social media mention), the density (i.e., the average number of social media metrics per publication), the trend analysis of publications mentioned in social media, and the correlation analysis of social media metrics with citation impact indicators received by these publications. All these four aspects (coverage, density, trend, and

[1] http://www.ref.ac.uk

[2] https://www.knaw.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021

correlations) are important in order to determine the scope and possibilities of social media metrics, as well as the potential (di)similarity with other bibliometric indicators (e.g., citations).

- *Coverage*

The results presented in Chapter 2 show that the social reference manager tool Mendeley is the most prevalent source of social media metric in terms of coverage of scientific publications across most disciplines. Overall, Mendeley is the social media source with the largest coverage of scientific publications (62.2%), followed by Twitter (1.5%), Wikipedia (1.4%), and Delicious (0.3%). In comparison to other reference managers such as CiteULike (Bar-Ilan et al., 2012; Priem, Groth, & Taraborelli, 2012) or BibSonomy (Haustein & Siebenlist, 2011) as well as also other social media platforms (Twitter, Facebook, Wikipedia, blogs, etc.), Mendeley has the highest coverage of publications (Haustein & Larivière, 2014; Priem, Piwowar, & Hemminger, 2012; Thelwall & Sud, 2016). However, the coverage of scientific publications in Mendeley varies across disciplines (see also Chapters 5 and 6). For example, more than 80% of publications in the Multidisciplinary sciences category of the Web of Science have some readership in Mendeley. Similarly, 73% of Medical and life sciences publications, 68% to 92% of s*ocial and behavioural sciences* publications (Chapters 1, 4, and 5), more than 90% of *Life and earth sciences* and *biomedical and health sciences*, of p*hysical sciences and engineering*, and 79% of publications from *mathematics and computer science* (Chapter 6) were covered in Mendeley. These results are in agreement with previous studies in which the coverage of publications from other databases (Scopus or PubMed) was studied. For instance, depending on the field, Mendeley covers 45% to 90% of the publications in the Scopus database (Thelwall & Sud, 2016) and more than 80% of publications published by PloS (Priem, Piwowar, & Hemminger, 2012). Some sub-fields of s*ocial sciences and humanities* (such as *sociology, communication, business, psychology, anthropology, educational research, and linguistics*) have a relatively high coverage and a relatively high number of readers in Mendeley (Costas, Zahedi, & Wouters, 2015; Hammarfelt, 2014; Mohammadi & Thelwall, 2014). The second most popular social media platform is Twitter but with a much lower coverage (20-30%) of publications than Mendeley (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013) and with a much stronger orientation towards the *social and medical sciences* (Costas, Zahedi, & Wouters, 2015). Other social media sources (such as Facebook, blogs, Google+, Reddit, LinkedIn, etc.) cover relatively much lower numbers of publications (less than 20% of PubMed or Web of Science) (Robinson-García et al., 2014; Thelwall et al., 2013).

- *Density*

In the *social sciences and humanities*, the density of Mendeley readership per publication is higher than the density of citations per publication (Chapters 4 and 5). For instance, in some sub-fields, particularly language, information & communication, and law, arts & humanities (Chapter 2) as well as *business, psychology, sociology, social and behavioral sciences,* *anthropology, education and educational research* (Chapter 5), the density of readership per publication outperforms that of citation counts. In contrast, publications from *mathematics and computer science* have the lowest readership and citation values. A recent study by Thelwall (2017) also showed that some subfields from *mathematics and computer science* (i.e., *analysis, algebra and number theory, geometry and topology*), *nuclear energy*, and *high energy physics* received on average the lowest levels of Mendeley readership of all fields. In comparison to other social media platforms the social media mentions from Twitter, Facebook, or blogs are also lower for papers from *mathematics and computer science* than those from *social sciences and humanities*, *biomedical and health sciences*, and *life and earth sciences* (Haustein, Costas, & Larivière, 2015). The variation in the density of readership could reflect the different uptake of Mendeley across different fields or the increasing popularity of other reference managers in some fields (e.g., preferring Zotero or BibSonomy over Mendeley). These results suggest the potential advantage of Mendeley readership over citations for the analysis of impact of publications particularly in the field of *social sciences*, which is a field that is not very well represented by citation databases (Nederhof, 2006).

- *Trends*

The trend analysis presented in Chapter 5 shows that the coverage of publications in Mendeley has increased from 84% in 2004 to 89% in 2009, and has decreased from 88% in 2010 to 82% in 2013. However, publications from 2010 onwards exhibit on average a higher density of readership scores than citation scores. This indicates that compared to citation scores, readership scores are more prevalent for recent publications. This result is in agreement with other studies which found that Mendeley readership scores are more prevalent in the most recent publication years than in the earliest years (Maflahi & Thelwall, 2016). Moreover, it is found that papers received Mendeley readership counts a year before they got cited (Thelwall & Sud, 2016) even if papers were not yet formally published (online first publication version) (Maflahi & Thelwall, 2018). The faster uptake and the stronger density of Mendeley reader counts for publications from the most recent years can be seen as a good proxy of "early scientific impact" for these publications (Thelwall & Sud, 2015). As citations need time to accumulate, they are less useful for evaluating recent publications. In contrast, readership can work as an early indicator of impact as they accumulate earlier and faster than citations (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011).

- *Relation with citation indicators*

In terms of relation with citations, a moderate spearman correlation ($r = 0.49$) has been found between Mendeley readership counts and citation indicators in this thesis (Chapter 2). A moderate correlation has been also found in other studies (Li & Thelwall, 2012; Maflahi & Thelwall, 2016; Mohammadi, Thelwall, Haustein, & Larivière, 2015). None of the other social media metrics exhibit a similar moderate correlation with citations (Costas et al., 2015a; Haustein et al., 2014). This suggests that Mendeley readership and citations are to some extent related activities.

Considering all the results presented in Chapter 2, the high coverage, density, and correlation of Mendeley readership with citations support the conclusion that readership counts capture a more scholarly type of impact, while other social media metrics such as Twitter, Facebook, or Wikipedia capture a more social media type of impact (Wouters et al., 2018). The latter is also reflected in the low coverage, density, and correlations of these metrics with citations (Costas, Zahedi, & Wouters, 2015b; Thelwall et al., 2013).

Q2: What are the most important challenges regarding data quality in the social media metrics offered by different altmetric data aggregators?

Chapter 3 provides answer to this question by studying the methodological choices used by the different altmetric data aggregators. This chapter also discusses how each altmetric data aggregator collects, processes, summarizes, and updates the social media metrics that they report. Main findings show that the same social media metrics collected for a same set of DOIs at the same time exhibit a substantial variability across different major altmetric aggregators. For instance, Lagotto and Plum Analytics provide the highest number of Mendeley readership as they aggregate the counts coming from different identifiers of the same paper or different forms of readership (e.g., individual readership and group readership). Altmetric.com provides the highest number of tweets, which can be explained by the tracking and combination of counts from different versions of the same object. Plum Analytics provides the highest number of Facebook counts as it combines different events from Facebook in the same score, and CrossRef Event Data provides the highest number of Wikipedia mentions, as it collects mentions from different languages and edits of the same Wikipedia entry. Similar results have been found in recent studies comparing different altmetric data aggregators (Meschede & Siebenlist, 2018; Ortega, 2017; Bar-Ilan and Halevi, 2017) as well as other previous studies (Chamberlain, 2013; Zahedi, Fenner, and Costas, 2014; 2015; Jobmann, et al., 2014).
The results of the correlation analysis also highlights that there are relevant differences in the social media metrics reported by different altmetric data aggregators. Mendeley readership counts exhibit the highest correlations, which means that the readership counts provided by all data aggregators are relatively consistent. The correlation analysis of Twitter counts also suggest a reasonably good agreement among data aggregators. In contrast, Facebook counts and Wikipedia counts have the lowest correlations among aggregators, caused by strong discrepancies in the Facebook/Wikipedia counts provided by each of these aggregators. The same high consistency across aggregators regarding Mendeley readership and similar levels of correlation between Mendeley readership, tweets, and Wikipedia mentions across similar altmetric data aggregators have also been highlighted in some previous studies (Meschede & Siebenlist, 2018; Ortega, 2017; Bar-Ilan and Halevi, 2017).

Based on these results, the most important challenges regarding data quality of social media metrics are formed by the following methodological choices adopted by the different altmetric data aggregators: *data collection choices*, *data aggregation and reporting choices*, and *updating choices*. These methodological choices affect the final counts and the conceptual meaning and interpretation of social media metrics provided by these aggregators. For instance, the choice of adding up different acts from the same social media source, like tweets or retweets, has conceptual repercussions, since a tweet can be seen as an act of greater engagement than a retweet (Haustein et al., 2016; Holmberg, 2015). Moreover, the adding up (or not) of different edits of the same Wikipedia entry has conceptual issues in the determination of the final Wikipedia impact of publications. The combination of conceptually different counts into one single metric may introduce misunderstandings, misuses, and even manipulations that could have negative effects on the further application of social media metrics. For instance, adding up tweets and retweets has conceptual repercussions since a tweet can be seen as an act of greater engagement than a retweet (Haustein, Bowman, & Costas, 2016; Holmberg, 2015). Hence, it seems reasonable to argue that keeping different events separate as much as possible and increasing the transparency of the methodological choices for the calculation and reporting of metrics is the best approach from an analytical perspective (Wouters et al., 2018). Therefore, based on the results of this chapter, altmetric data aggregators should increase the transparency of their methodological choices in data collection, aggregation, and calculation of their metrics. Altmetric data users, researchers, and data aggregators should be aware of the unintended effects that these methodological choices can have in the valid use and application of social media metrics data. Understanding how methodological and technical choices can influence the analytical reliability and validity of social media metrics is a crucial element in the future development of the social media studies of science.

Q3: What are the main characteristics of publications saved and read on Mendeley?

As shown in this thesis, Mendeley readership is one of the most prominent social media metric sources, with a stronger scholarly orientation compared to other social media metrics. Hence, the question in Chapter 4 is what kind of publications are being saved in Mendeley. Chapter 4 answers this question by studying the relationship between typical bibliographic document characteristics and citations with Mendeley readership.

The findings reveal that document types like editorial materials, letters, news items, book reviews or meeting abstracts have a much higher coverage in Mendeley as well as a much higher readership density than citations. These document types focus more on disseminating scientific debates, news, opinions, or summarized information, and typically receive relatively less citations. Due to their lower citation density, they are deemed not suitable for robust citation analysis and are often excluded from citation analyses (Waltman et al., 2011).

Publications with relatively higher Mendeley readership counts are also related to the same bibliographic characteristics as those observed for publications with relatively higher citation counts. For instance, collaborative papers and papers with more references are more frequently saved in Mendeley, which is similar to the higher citation rates received by papers with the same characteristics. The distribution of citations and readership across disciplines exhibit remarkably similar patterns of skewness (Costas, Haustein, Zahedi, & Larivière, 2016; Costas, Perianes-Rodríguez, & Ruiz-Castillo, 2017). The strong similarities between citations and readership have paved the way for the development of field-normalized readership indicators (Haunschild & Bornmann, 2016).

These results, reinforce the idea that Mendeley readership and citations are two different but connected processes capturing a similar type of impact. The moderate correlations found between citation and readership counts (Thelwall & Wilson, 2015; Costas et al., 2015; Haustein et al., 2014b; Zahedi et al., 2014) and the “pre-citation role” that is attributed to Mendeley readership (i.e., that Mendeley users save documents in their libraries to cite them later (Haustein, Bowman & Costas, 2015; Thelwall & Sud, 2015)) reinforce the idea that these two indicators are very similar, both conceptually and empirically. However, the existence of two indicators related but not equivalent, that capture a similar concept and that can be used for the analysis of different academic actors opens the debate on how they should be interpreted when divergent results are provided by each one (e.g., a hypothetical case in which a University is low on citations and high on readership compared to another one high on citations and low on readership) as suggested by Costas, Perianes-Rodríguez, & Ruiz-Castillo, (2017). All this clearly points to the need of further studies in order to better understand the differences, similarities and complementarities between these two metrics (i.e., citations and readership).

Q4: What are the practical analytical possibilities of Mendeley readership metrics?

This question is answered through two sub-questions presented in Chapters 5 and 6.

Chapter 5 discusses whether Mendeley readership would be more useful than journal-based indicators in identifying highly cited publications. The identification of highly cited publications is a critical element in bibliometric research as well as in research evaluation (Aksnes, 2003; Ivanović & Ho, 2014). Highly citedness can be a sign of the quality, relevance, or scientific excellence of papers or even an indicator of breakthrough research (Bornmann, 2014; Schneider & Costas, 2014). Although highly citedness doesn’t always reflect the research quality of publications (Waltman, van Eck, & Wouters, 2013), it is considered as a relevant indicator in research evaluation in a large number of fields (Abramo & D’Angelo, 2015; Tijssen, Visser, & van Leeuwen, 2002). Using journal-level indicators in identifying high quality publications, researchers, or research groups has been a common practice in research evaluation (Rushforth & Rijcke, 2015; Jiménez-Contreras, de Moya Anegón, & López-Cózar, 2003). However, as journal indicators are usually considered bad proxies of the impact of individual publication (although critiques of this idea have been voiced recently, cf. Waltman & Traag, 2017), their use in evaluating individual publications has been widely debated in the literature (Adler, Ewing, & Taylor, 2008; Larivière et al., 2016; Seglen, 1997). Some initiatives such as DORA[3] and the Leiden manifesto[4] have also criticized the use of journal-based indicators for the analysis of individual publications. In the context of this debate, whether or not social media metrics are better able to identify highly cited publications over journal-based indicators gains importance (Waltman & Costas, 2015).

Thus the first sub-question was *whether Mendeley readership scores can identify highly cited publications more effectively than journal citation scores*. Chapter 5 answers this sub-question. It is demonstrated that Mendeley readership counts are indeed more effective (in terms of precision/recall values) than journal-based indicators in filtering highly cited publications across all fields of science and publication years. This is in contrast to other social media metrics (e.g., F1000 recommendations, Twitter, blogs, and Facebook counts) that have not been found to have such a property (Waltman & Costas, 2014; Costas, Zahedi, & Wouters, 2015a). Therefore, the result of this chapter shows for the first time a practical advantage of a social media metric (readership counts) over another more established bibliometric indicator (e.g., the Journal Impact Factor), and opens the door to incorporating Mendeley readership as a valid and relevant indicator for the prediction of future citations (Zahedi, Costas, & Wouters, 2017). Hence, it can be concluded that Mendeley readership scores are an effective tool to filter highly cited publications and it can indeed play a role as an alternative approach (to journal-based impact indicators) to find highly-cited outputs.

A distinctive feature of Mendeley readership counts is that they can be broken down by types of users of Mendeley (e.g., Master students, PhDs, Professors, etc.). Although the number of Mendeley readership counts do not necessarily reflect the actual reading of publications (Haustein et al., 2016; Mohammadi et al., 2014), Mendeley readership counts can be used for the identification of the scientific, educational, or professional interests of different publications and disciplines based on their reception by different Mendeley user types (Zahedi & Van Eck, 2018). Hence, the analysis of the readership counts of scholarly publications of Mendeley users enables the analysis of different forms of reception (or impact) of scientific publications. Chapter 6 studies the mapping of disciplinary differences in readership counts by types of Mendeley users (e.g., professors, researchers, students, professionals, and librarians). The sub-question in this chapter is: *what are the topics of interest of different Mendeley users and how do their use of scholarly documents reflect different types of impact of research.* The results of this chapter show that the largest uptake of Mendeley is by students. Also, professors and students are mainly active in the *social sciences & humanities* and *mathematics & computer science*, which are fields with a low citation density in the Web

[3] https://sfdora.org/
[4] http://www.leidenmanifesto.org/

of Science (Nederhof, 2006). In contrast, researchers and other professionals are mainly active in fields with a high citation density such as *life & earth sciences*, *physical science & engineering*, and *biomedical & health sciences*. These results are in line with other studies (Haustein & Larivière, 2014; Mohammadi, Thelwall, Haustein, and Larivier, 2015, Zahedi & Van Eck, 2015; Thelwall, 2017) that show that substantial differences in readership practices exist between (sub)fields and user types. In addition, in terms of topics of interest, the results of this chapter indicate that user types pay more attention to publications related to their roles and the purpose for which they use Mendeley. For instance, professors mostly save publications related to teaching and educational topics (e.g., higher education, medical education, and second language acquisition). This may be expected since professors use Mendeley among other reasons to organize literature for teaching and publishing. Fundamental or theoretical papers (such as business, management, and leadership) as source of course materials or as a source of reading for thesis work are more interesting for students. Researchers are relatively more interested in research fronts and applied sciences (climate research, pharmaceutical research, and biotechnology to astronomy and astrophysics). These results are in line with another study that showed that F1000 publications with the tag 'good for teaching' (papers with a good overview of a topic) were more relevant for Mendeley users classified as lecturers, while papers with the tag 'new findings' were mostly read by researchers (Bornmann & Haunschild, 2015). Librarians show relatively more interest in topics related to their work, namely bibliometrics and scientometrics, library science, research utilization, and clinical guidelines. Not surprisingly, the user group professionals (which includes for example medical doctors, nutritionists, and lawyers) is relatively more interested in practical and technical oriented topics (e.g., biological, medical, and clinical oriented topics). These results show that publications saved by different user types can be related to different contexts of use, such as education, (self) training, research, or practical and applied uses. Thus, although there is not enough evidence in the literature (Thelwall, 2016), publications mostly saved by students can be seen to have an *educational interest*, those saved by professionals to have a more *professional interest*, and the ones saved mostly by professors or researchers can be related to more scientific interests. The results in Chapter 6 emphasize the potential role of readership indicators for capturing the usage of scientific documents by a wide range of audiences.

7.3. Answer to the main question of this thesis

The main question of this PhD thesis is *what is the potential usefulness and added value of social media metrics for informing research evaluation?*.

The different results presented in this thesis demonstrate that Mendeley readership is the social media metric source with the strongest usefulness and added value for research evaluation. This is justified based on the large coverage, density, correlation, document characteristics, and conceptual proximity of Mendeley readership with citation indicators. This becomes specially clear when compared to other social media metrics (e.g., Twitter, Facebook, or Wikipedia counts) with a more marginal coverage and density, much lower correlation, and fundamental conceptual differences with citations.

In addition to the above, some specific characteristics of Mendeley readership also support the added value of this source for research evaluation. Thus, Mendeley readership counts can be more valuable to inform the evaluation in fields like *social sciences and humanities*, which are typically not well represented with citation indicators (Chapter 2). Also for some document types (e.g., editorial material, letters, etc.) which are typically excluded from citation analysis, Mendeley readership indicators can play a relevant role for their analysis (Chapter 4). Moreover, they can be used for the identification of highly cited publications as an alternative to journal-based indicators (Chapter 5). The analysis of readership patterns by different user types (students, professors, professionals, etc.) has also been proved as a new source for studying different forms of reception and impact of application, thus expanding beyond the more academic impact captured by citations (Chapter 6).

Regarding other social media metrics such as Twitter and Facebook mentions, Wikipedia citations, etc. the results of this PhD thesis do not support the use of these other metrics in the same way citations (and also Mendeley readership) are used in research evaluation. This is justified by their lower coverage, correlations with citations as well as their very different conceptual features (e.g., lack of an academic orientation, free nature, gaming possibilities, etc.). This leaves open the question of what role they could play for research evaluation, if any. In a recent publication (Wouters, Zahedi, Costas, 2018) it has been proposed that these more social media focused metrics, precisely given their lower scholarly orientation while having a higher social media orientation, could still play a role in the evaluation of social media-related activities, thus allowing the evaluation of the reception of scholarly outputs among social media users, the spread of ideas, or the presence of academic entities (e.g., scholars or universities) on social media platforms.

7.4. Implication of the findings of this thesis

The results of this PhD thesis imply that social media metrics are important sources of information about the saving, commenting, sharing, and discussing of scientific publications by different audiences across multiple social media platforms. The diverse and user-generated social media metrics provide a broad overview of how, when, from where, by whom, and with whom scientific publications are mentioned, saved, shared, and discussed. Hence, due to their advantages such as their speed, openness, and diversity (Wouters & Costas, 2012), social media metrics provide new possibilities for measuring research performance. This information together with other indicators of usage of scientific publications such as downloads, views, and citations could extend the concept of *impact* of research.

Based on the results of this thesis and given the importance of social media in scholarly communication, some uses of social media metrics, particularly readership indicators, are suggested to inform the process of research evaluation. The advantages of social media metrics of being immediately available and reflecting wider audiences of scholarly publications make their considerations in research evaluation relevant (Wouters & Costas, 2012). As social media metrics are timely indicators of various user's engagement with diverse scholarly objects, these indicators can reveal early impact of research, various types of engagement of user community with research objects, and different aspects of impact of research. These features extend the concept of research impact beyond any specific actors, databases, and geographic limitations and enable their use in informing decisions regarding research evaluation. However, due to the social media nature of these metrics they are susceptible to easily gaming and moreover changing fast. In addition, data quality and transparency challenges, potential biases in the visibility of research from different disciplines across geographic landscape as well as inequality in access, availability, and use of social media sources across different nations challenge their reliability as research evaluation indicators. Hence, depending on the unit of analysis (individual publication, researcher, research group, university, or country levels) relevant questions should be asked before interpreting the results based on these metrics, including:

- *to what extent are publications from specific discipline and publication years of a given unit covered by the social media platforms under analysis?*
- *to what extent are the individuals or the publications from the same institute, research group, university or the country under the analysis visible or represented on the social media platforms?*
- do *the unit of analysis use social media platforms and if so for what purposes?*
- *to what extent are the social media platforms which is the source of metrics known, accessible, and in use by users from different countries?*
- And finally *what are the data quality and limitations of the metrics provided by the platforms?*

From a more conceptual point of view, the results of this PhD thesis support the framing of Mendeley counts as proxies of the *intention of reading* of their users. Readership counts can be expected to capture a relatively low level form of engagement of interaction between the users and their publications (Haustein et al., 2016)[5], in which the act of saving documents in Mendeley can be considered as a basic signal of the users potentially having interest in reading them at some point (although they may actually end up never reading them). This basic framing legitimates the use of readership data for the analysis of the reading interest of scientific publications, which can be useful not only for evaluative purpose (as demonstrated in this PhD thesis), but also for the development of library collections, reading guides, or reading recommendation systems. More importantly, this concept of readership as a token of the *intention of reading* also suggests the existence of broader frameworks, in which other reading-related processes (such as opening publications, scrolling, actual reading, highlighting

[5] This argument is justified since current Mendeley readership counts do not capture the actual act of reading (as it is not possible to know whether the users have actually read them).

parts of the text, writing notes in the text, assigning tags, reading frequency and time spent, commenting, citing, etc.) could be captured. Information about these processes would inform, in a more advanced manner, how the different users are engaging with scientific content. The development of such advanced frameworks, which is beyond the scope of this PhD thesis, could provide more advanced insights into the reading behaviour of users, eventually allowing the development of more advanced indicators and a richer perspective on the interest and reception of publications by their users. These ideas about future research are further expanded in the next section.

7.5. Further research prospects

The work presented in this PhD thesis opens several new paths of further research. In this section we summarize the most important ones. As it has been shown, Mendeley readership is one of the most promising sources of social media metrics; therefore, our further research lines will focus mostly on readership indicators, although some of the lines suggested here could be also considered for any other social media metrics source. These research possibilities can be organized in three major areas of further development: *improvement and expansion of available readership data*, *development of new readership indicators and analytics*, and *further understanding and theorization of readership*.

Improvement and expansion of available readership data

The research developed in this thesis has been bounded by multiple data limitations imposed by most altmetric data aggregators, and particularly those imposed by Mendeley. Future research on readership would indeed benefit from incorporating additional data elements currently not available. Some elements that could be included in the future analysis of Mendeley readership include the *timelines of readership* (i.e., studying when publications are saved -or read- by each user in the system). This would allow the determination of readership windows, the study of the accumulation of readership over time, and the temporal analysis of readership. This is important in order to be able to study readership impact considering fixed or variable windows as it is currently done for citations (Abramo, Cicero, & D'Angelo, 2011). Currently, the only possibility to study Mendeley readership is using variable citation windows (i.e., until the moment of collecting the readership data).

Another interesting element is the availability of information on the *deletion of publications* from users' libraries. This is a quite distinctive element of readership information not existing for citations (since citations once given become permanent) that could provide unique information about information obsolescence and relevance. Tracking the trends of addition and deletion of publications to and from users libraries reveals the relevance and outdating of those publications for those users. This information can be informative in different contexts. For instance for decisions making regarding the relevance of information materials in the libraries and information centers based on the usefulness of information overtime as well as knowledge accumulation and obsolescence.

Other quite distinctive characteristic of readership is that each readership event is provided by one single user. This information enables more direct *user-publication relationship studies*, in which it is possible to determine and study the engagement, usage, and interest of a specific user on a given publication. This type of analysis is important particularly when put in contrast to citations. In papers with multiple authors it is not possible to discern which author has included which citations, being impossible to attribute what was the interest or relevance of specific publications for individual scholars. The use of Mendeley readership counts allows to better determine the usage, interest, and interaction of specific users for specific sets of publications, thus allowing to answer the question of *who is interested in what?* This information opens the possibility of developing more focused information behavior studies of groups of users. This type of studies would be only possible if Mendeley would provide more individualized data on the saving and usage patterns of individual users.

In general, larger availability of *user data* would be a fundamental element in order to further study readership behaviours, including aspects related with users' age, gender, disciplines, academic status, affiliations, or priority for different reading items. This information, which is still not disclosed, would allow the expansion of the research agenda towards more demographic and sociologic aspects of readership patterns, provided of course that users' privacy rights are respected and protected. Some altmetric data aggregators (such as Altmetric.com[6]) have already updated their privacy policy regarding the collection and recoding of personal information based on the new EU General Data Protection Regulation (GDPR)[7]. However, since readership data obtained from Mendeley is anonymous disclosing the above information (users' age, gender, etc.) would not be against the new data protection regulation. In line with the above, the expansion and development of information about the users would also allow the development of indicators of engagement and appraisal of publications. Thus, indicators on how many times a publication has been opened by a user, for how long, whether the user has scrolled through the publication, assigned tags to the items, written notes, highlighted parts of the text, etc. would be possible. The aim is to develop indicators of *reading and appraisal* able to capture the real interaction and engagement of the users with the publications. Thus, it would be possible to identify how users of the reference managers value and appraise the publications they have saved in their libraries. In fact, as explained above, depending on the evaluative context such type of indicators could be even more meaningful and informative than citations since they would say something about the actual value and consideration of the publication by the users. This is something that is not possible with citations, which tend to be more neutral and may often be of a perfunctory nature (Waltman et al., 2013).

The above elements need to be considered with the existence of multiple online reference managers other than Mendeley (e.g., Zotero, CiteULike, BibSonomy in case their data become

[6] https://help.altmetric.com/support/solutions/articles/6000196080-gearing-up-for-gdpr
https://www.altmetric.com/privacy-policy/

[7] General Data Protection Regulation (GDPR): https://ec.europa.eu/info/law/law-topic/data-protection_en

available through open APIs) that can also provide information on the interactions and engagement between users (readers) and publications. Therefore it will be also important to develop research on *the coverage, data issues, complementarities, and usefulness* of these online reference managers to reflect types of use, appraisal, intentions, and engagement of users with scholarly publications. This would provide a broader perspective on readership habits.

Development of new readership indicators and analytics

An important question that will need to be addressed in the near future is how readership can specifically be valued in the context of research evaluation and scientometric work. Thus questions around *what is the value of readership scores* for science policy makers, research actors and stakeholders and how these values can be incorporated in the actual evaluation of science will need to be addressed in future work. This PhD thesis has paved the way, by demonstrating how Mendeley readership can be valuable in specific fields (e.g., *social science and humanities*) which are typically not well served by citation indicators. This also holds for some document types (e.g., editorial material, letters, etc.), which also are not well represented by citation indicators. The analysis of readership patterns by different user types (students, professors, professionals, etc.) is a new source for studying forms of reception and impact beyond the more academic impact captured by citations. Moreover, the possibility to track patterns of use (saving or reading) of scientific outputs in different languages and from different countries enables the study of technological, cultural, and political factors that could affect the social media reception of publications across nations and cultures, helping to identify potential biases and the so-called *altmetric divide* (understood as the inequality in the access and use of social media platforms across different countries, which leads to biases in the social media metrics from different countries) (Zahedi, 2016; Zahedi & Costas, 2017).

Not only further research on more advanced readership indicators (e.g., based on the engagement of the users with the publications as mentioned above) will be important, but also new *network-based indicators* like readership coupling or co-readership as recently suggested (Costas, De Rijcke, & Marres, 2017; Kraker, Schlögl, Jack, & Lindstaedt, 2015) would deserve further attention. By mimicking the network analysis of citations and bibliometric sources, these network-based indicators would allow the clustering of users (or readers) by their common topics of interest as well as studying how these users are connecting scientific topics by their readership habits.

Finally, research on the potential *prediction* of indicators based on readership and social media data is an important future topic that will have to deal with conceptual and theoretical issues on the relationships and dependencies among indicators. Thus, how tweeting a publication may be related to that publication being blogged, and this itself being related to the publication being saved on Mendeley or eventually becoming highly cited are all aspects that will deserve more attention in the future.

Further understanding and theorization of readership

The reinforcement and expansion of the theoretical foundations of social media metrics is still needed. Thus, the development of theoretical frameworks concerning the relations between citations and bibliometrics indicators, readership indicators and other social media metrics will be necessary in the near future. The combination of different theories coming from different disciplines (e.g., science of science, science and technology studies, sociology of science, citation theories, social media theories, etc.) (Haustein et al., 2016) will be necessary in order to develop comprehensive *readership theories* that will help to understand why people read what they read, cite what they cite, tweet what they tweet, and how all these acts relate to each other. For example, it will be important to further explore the relation between getting mentioned in one social media platform and how this can influence the cycle of social media mentions and reception across other different platforms. Results presented in this PhD thesis have made clear how the study of readership will be an important component in the further development of the broader *social media studies of science* (Costas, 2017; Wouters et al., 2018). From the perspective of the social media studies of science, the interactions between social media actors (e.g., Mendeley users, tweeters, bloggers) and scholarly entities (e.g., publications, scholars, academic organizations, scientific journals, etc.) will become the focal point, helping to expand our understanding of the influences that these two realms (social media and science) are having on each other.

7.6. References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics*, *5*(4), 659–667. http://doi.org/10.1016/j.joi.2011.06.004

Abramo, G., & D'Angelo, C. A. (2015). Ranking research institutions by the number of highly-cited articles per scientist. *Journal of Informetrics*, *9*(4), 915–923. http://doi.org/10.1016/j.joi.2015.09.001

Adler, R., Ewing, J., & Taylor, P. (2008). *Citation statistics. A report from the International Mathematical Union.* Retrieved from:
www.mathunion.org/publications/report/citationstatistics0.

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, *12*(3), 159–170. http://doi.org/10.3152/147154403781776645

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. Retrieved from:
http://arxiv.org/abs/1205.5611

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, *23*(2), 166–173. http://doi.org/10.1093/reseval/rvu002

Bornmann, L., & Haunschild, R. (2015). Which people use which scientific papers? An evaluation of data from F1000: Discovery Service para Universidad de La Rioja. *Journal of Informetrics*, *9*(3), 477–487. http://doi.org/http://dx.doi.org/10.1016/j.joi.2015.04.001

Costas, R. (2017). Towards the social media studies of science: social media metrics, present and future. *Bibliotecas. Anales de Investigación*, *13*(1), 1–5.

Costas, R., De Rijcke, S., & Marres, N. (2017). Beyond the dependencies of altmetrics: conceptualizing "heterogeneous couplings" between social media and science. In *altmetrics17. The dependencies of altmetrics*. Toronto, Canada.

Costas, R., Haustein, S., Zahedi, Z., & Larivière. (2016). Exploring paths for the normalization of altmetrics: applying the Characteristic Scores and Scales. In *altmetrics16. Moving beyond counts: integrating context* (pp. 1–8). Bucharest, Romania.

Costas, R., Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2017). On the quest for currencies of science: field " exchange rates " for citations and Mendeley readership. *Aslib Journal of Information Management*. http://doi.org/10.1108/ AJIM-01-2017-0023

Costas, R., Zahedi, Z., & Wouters, P. (2015a). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. http://doi.org/10.1002/asi.23309

Costas, R., Zahedi, Z., & Wouters, P. (2015b). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. http://doi.org/10.1002/asi.23309

Costas, R., Zahedi, Z., & Wouters, P. (2015c). The thematic orientation of publications mentioned on social media Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, *67*(3), 260–288. http://doi.org/10.1108/AJIM-12-2014-0173

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, *101*(2), 1419–1430. http://doi.org/10.1007/s11192-014-1261-3

Haunschild, R., & Bornmann, L. (2016). Normalization of Mendeley reader counts for impact assessment. *Journal of Informetrics*, *10*(1), 62–73.
http://doi.org/10.1016/j.joi.2015.11.003

Haustein, Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2013). Coverage and adoption of altmetrics sources in the bibliometric community. In *Proceedings of the 14th International Society of Scientometrics and Informatics Conference,* 15-19th July 2013, Vienna, Austria.

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting Altmetrics: Viewing Acts on Social Media through the Lens of Citation and Social Theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication* (pp. 372–406). Berlin, Boston: De Gruyter. http://doi.org/10.1515/9783110308464-022

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE*, *10*(3), e0120495. http://doi.org/10.1371/journal.pone.0120495

Haustein, S., & Larivière, V. (2014). Mendeley as a source of global readership by students and postdocs? Evaluating article usage by academic status. In *In 35th International Association of Technological University Libraries (IATUL), 2-5 June 2014, Aalto University, Helsinki, Finland. Helsinki* (pp. 1–10). Helsinki, Finland.

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *Information Technology*, *56*(5), 207–215. http://doi.org/10.1515/itit-2014-1048

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*. http://doi.org/10.1016/j.joi.2011.04.002

Holmberg, K. (2015). Classifying altmetrics by level of impact. In *Proceedings of the 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey.*

Ivanović, D., & Ho, Y. S. (2014). Highly cited articles in the Information Science and Library Science category in Social Science Citation Index: A bibliometric analysis. *Journal of Librarianship and Information Science*. http://doi.org/10.1177/0961000614537514.

Jiménez-Contreras, E., de Moya Anegón, F., & López-Cózar, E. D. (2003). The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, *32*(1), 123–142.

Kraker, P., Schlögl, C., Jack, K., & Lindstaedt, S. (2015). Visualization of co-readership patterns from an online reference management system. *Journal of Informetrics*, *9*(1), 169–182. http://doi.org/10.1016/j.joi.2014.12.003

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., … Curry, S. (2016). A simple proposal for the publication of journal citation distributions. *A Simple Proposal for the Publication of Journal Citation Distributions.*, *62109*. http://doi.org/10.1101/062109

Li, X., & Thelwall, M. (2012). F1000, Mendeley and Traditional Bibliometric Indicators. *Proceedings of the 17th International Conference on Science and Technology Indicators*, *3*, 451–551.

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. http://doi.org/10.1002/asi.23369

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. *Scientometrics*. http://doi.org/10.1007/s11192-018-2674-1

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.

http://doi.org/10.1002/asi.23286

Mohammadi, E., Thelwall, M., Kousha, K., Cybermetrics, S., Science, C., & Street, W. (2014). Can Mendeley Bookmarks Reflect Readership ? A Survey of User Literature review Changes in scholarly reading habits in the digital era. *Journal of the Association for Information Science and Technology*, 67 (5).

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, *66*(1), 81–100.

http://doi.org/10.1007/s11192-006-0007-2

Ortega, J. L. (2017). Reliability and accuracy of altmetric providers : a comparison among Altmetric, PlumX and Crossref Event Data. Retrieved from:

https://www.researchgate.net/publication/319457093_Reliability_and_accuracy_of_altmetric_providers_a_comparison_among_Altmetric_PlumX_and_Crossref_Event_Data?discoverMore=1

Priem, J., Groth, P., & Taraborelli, D. (2012). The Altmetrics Collection. *PLoS ONE*, *7*(11), e48753. http://doi.org/10.1371/journal.pone.0048753

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved from http://arxiv.org/abs/1203.4745

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). altmetrics : a manifesto. Retrieved from http://altmetrics.org/manifesto/

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R. (2014). New Data, New Possibilities: Exploring the Insides of Altmetric.Com. *Nuevos Datos, Nuevas Posibilidades: Revelando El Interior de Altmetric.com.*, *23*(4), 359–366. http://doi.org/10.3145/epi.2014.jul.03

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., Costas, R., Robinson-García, N., Torres-Salinas, D., … Costas, R. (2014). New Data, New Possibilities: Exploring the Insides of Altmetric.Com. *El Profesional de La Information*, *23*(4), 359–366. http://doi.org/10.3145/epi.2014.jul.03

Schneider, J. W., & Costas, R. (2014). Identifying potential “breakthrough” research articles using refined citation analyses: Three explorative approaches. In E. Noyons (Ed.), *Proceedings of the 19th Science and Technology Indicators Conference, 3-5 September, Leiden University, The Netherlands.*

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal(BMJ)*, *314*, 498–502.

Thelwall, M. (2016). Does Mendeley provide evidence of the educational value of journal articles? *Learned Publishing*, (August), 1–7. http://doi.org/10.1002/leap.1076

Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, *113*(3), 1721–1731. http://doi.org/10.1007/s11192-017-2557-x

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, *8*(5), e64841.
http://doi.org/10.1371/journal.pone.0064841

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. http://doi.org/10.1002/asi.23559

Tijssen, R. J. W., Visser, M. S., & van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, *54*(3), 381–397. http://doi.org/10.1023/A:1016082432660

Waltman, L., & Traag, V. A. (2017). Use of the journal impact factor for assessing individual articles need not be wrong, 1–32. Retrieved from http://arxiv.org/abs/1703.02334

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, *87*(3), 467–481. http://doi.org/10.1007/s11192-011-0354-5

Waltman, L., van Eck, N. J., & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of Informetrics*, *7*(3), 635–641.
http://doi.org/10.1016/j.joi.2013.04.001

Wouters, P., & Costas, R. (2012). Users , narcissism and control – tracking the impact of scholarly publications in the 21 st century. Utrecht: SURFfoundation. Retrieved from http://www.surf.nl/nl/publicaties/Documents/Users narcissism and control.pdf

Wouters, P., Zahedi, Z., & Costas, R. (2018). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of Quantitative Science and Technology Research*. Springer.

Zahedi, Z., & Costas, R. (2017). How visible are the research of different countries on WoS and twitter? An analysis of global vs. local reach of WoS publications on Twitter. In *proceedings of the 16th International Conference on Scientometrics and Informetrics, Wuhan, China.*

Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, *68*(10), 2511–2521. http://doi.org/10.1002/asi.23883

Zahedi, Z., & Van Eck, N. (2015). Identifying topics of interest of Mendeley users using the text mining and overlay visualization functionality of VOS viewer. In *20th International Conference in Science & Technology Indicators, 2-4 September 2015, Lugano, Switzerland.*

Zahedi, Z., & Van Eck, N. J. (Accepted). An analysis of topics of interest of Mendeley users using VOSviewer overlay visualization approach. *Journal of Altmetrics*.

APPENDICES

- Summary
- Nederlandse Samenvatting
- Supplementary information to Chapters 3, 4, 5, & 6
- List of publications
- Acknowledgement
- Curriculum Vitae

Summary

The availability of indicators based on social media has opened the possibility to track the online interactions between social media users and scholarly entities. Indicators derived from these online interactions reflect aspects such as *how often*, by *whom*, and *when* are scholarly publications mentioned and discussed on social media platforms. These new indicators, popularly known as *altmetrics* and more technically referred as *social media metrics* are usually proposed as potential alternatives to citation-based indicators to inform research evaluation. The main aim of this PhD thesis is to explore the possibilities of social media metrics for informing research evaluation. The main ambition is to increase the knowledge and understanding of the limitations, challenges, and actual possibilities of social media metrics for research evaluation. The main research question that this thesis addresses is *what is the potential usefulness and added value of social media metrics for informing research evaluation?*. The research presented in this PhD thesis provides both empirical and conceptual answers for the consideration of social media metrics in research evaluation. The thesis is structured in seven chapters.

Chapter 1 presents a general introduction to social media and scholarly communication. It discusses the origins, definitions, and data availability (through different altmetric data aggregators) of social media metrics. It reviews the challenges, limitations, and possibilities of social media metrics for research evaluation. It describes different social media metrics data sources, particularly focusing on Mendeley as a specific relevant data source for research evaluation. Finally, this chapter introduces the main aim and research questions of this thesis.

Chapter 2 provides a general overview of the presence and coverage of publications presented in social media platforms and the distribution of social media metrics across fields, publication years, and document types. This chapter gives some important insights into the extent to which scientific publications are covered across social media platforms, the amount of social media attention received by them, and disciplinary differences in their social media metrics reception. This chapter also describes the relationship between social media metrics and citation indicators. Considering all the results presented in Chapter 2, the high coverage, density, and correlation of Mendeley readership with citations support the conclusion that readership indicators capture a more scholarly type of impact, while other social media metrics such as Twitter, Facebook, or Wikipedia capture a more social media type of impact.

Chapter 3 provides a thorough analysis of the most important data quality challenges and issues regarding social media data provided by the major altmetric data aggregators. This chapter discusses how the data collection and reporting approaches of these altmetric data aggregators influence both technically and conceptually the metrics provided. Main findings show that the same social media metrics collected for a same set of DOIs at the same time exhibit a substantial variability across different major altmetric aggregators. The most

important challenges regarding data quality of social media metrics can be related to the following methodological choices adopted by the different altmetric data aggregators: *data collection choices*, *data aggregation and reporting choices*, and *updating choices*. Based on the results of this chapter, some recommendations for altmetric data aggregators are put forward. These recommendations include increasing the transparency around the methodological choices in data collection, aggregation, and calculation of their metrics by the altmetric data aggregators. Altmetric data users, researchers, and data aggregators should be aware of the unintended effects that these methodological choices can have in the valid use and application of social media metrics data. Understanding how methodological and technical choices can influence the analytical reliability and validity of social media metrics is a crucial element in the future development of the social media studies of science.

Chapter 4 describes the disciplinary differences in the relationship between Mendeley readership and citation counts with particular documents' bibliographic characteristics across a dataset of 1.3 million publications from the Web of Science. The association between Mendeley readership, citation counts, and document characteristics (i.e., document types, number of pages, length of titles, length of reference lists, number of authors, institutes and countries) has been investigated using Ordinary Least Square (OLS) regression analysis. The findings reveal that document types like editorial materials, letters, news items, book reviews or meeting abstract have a much higher coverage in Mendeley as well as a much higher readership density than citations. Publications with relatively higher Mendeley readership counts are also related to the same bibliographic characteristics as those observed for publications with relatively higher citation counts. The chapter contributes to the identification of document-related differences between Mendeley readership and citations. This information is useful for the future construction of appropriate and meaningful indicators based on Mendeley readership.

Chapter 5 presents the results of a large-scale analysis of the distribution and presence of Mendeley readership scores over time and across disciplines across 9.1 million publications from Web of Science from the years 2004-2013. The results of this chapter show that Mendeley readership counts are indeed more effective (in terms of precision/recall values) than journal-based indicators in filtering highly cited publications across all fields of science and publication years. It is concluded that Mendeley readership indicators are a more effective tool to filter highly cited publications than journal-based citation impact indicators. This conclusion opens the door to incorporating Mendeley readership as a valid and relevant indicator for the prediction of future citations.

Chapter 6 focuses on the different user types in Mendeley and their thematic orientations. A dataset of 1.1 million Web of Science publications from the year 2012 are analyzed. The disciplinary differences in the reading (saving) patterns of different Mendeley user types are depicted using VOSviewer maps. Topics of interest of different user types in Mendeley are analyzed. The results of this chapter indicate that different user types have relatively more attention for publications related to their roles and the purpose for which they use Mendeley. The results point to the idea that publications saved by different user types can be related to different contexts of use, such as education, (self) training, research, or practical and applied uses. These results suggest that the analysis of the readership by different Mendeley user types can be used for the identification of the *scientific*, *educational*, or *professional* interests of different sets of publications. The results in Chapter 6 also emphasize the potential role of readership indicators for capturing the usage of scientific documents by a wide range of audiences.

Finally, **Chapter 7** includes the discussion and conclusions of the main results of this PhD thesis. It presents the summary of findings and the implications of the results obtained for informing research evaluation, together with some perspectives for further research. The different results presented in this thesis clearly demonstrate that Mendeley readership is the social media metric source with the strongest usefulness and added value for research evaluation. This is justified based on the large coverage, density, correlation, document characteristics, and conceptual proximity of Mendeley readership with citation indicators. This stronger added value of Mendeley readership for research evaluation becomes specially clear when compared to other social media metrics (e.g., Twitter, Facebook, or Wikipedia counts) with a more marginal coverage and density, much lower correlation, and more fundamental conceptual differences with citations. Some possible directions for further research based on the result of current work are presented and discussed, particularly in the direction of developing more advanced readership indicators (e.g., by incorporating more detailed information on the Mendeley users' interactions with scholarly outputs) together with the development of *readership theories* to better understand the behavior of online readers.

Nederlandse Samenvatting

De beschikbaarheid van sociale media-indicatoren heeft het mogelijk gemaakt de online interacties tussen gebruikers van sociale media en wetenschappelijke entiteiten te volgen. Indicatoren van deze online interacties weerspiegelen aspecten zoals hoe vaak, door wie en wanneer wetenschappelijke publicaties worden genoemd en besproken op sociale mediaplatforms. Deze nieuwe indicatoren, in de volksmond beter bekend als altmetrics en meer technisch aangeduid als social media metrics, worden vaak voorgesteld als mogelijke alternatieven voor citatie-indicatoren om onderzoeksevaluatie te ondersteunen. Het doel van dit proefschrift is om de mogelijkheden van sociale media voor onderzoeksevaluatie te verkennen. De ambitie is om de kennis en het begrip van de beperkingen, uitdagingen en feitelijke mogelijkheden social media metrics voor onderzoeksevaluatie te vergroten. De belangrijkste onderzoeksvraag van dit proefschrift is wat de potentiële bruikbaarheid en toegevoegde waarde van social media metrics voor onderzoeksevaluatie is. Het onderzoek gepresenteerd in dit proefschrift biedt zowel empirische als conceptuele antwoorden voor het gebruik van social media metrics in onderzoeksevaluatie. Het proefschrift is verdeeld in zeven hoofdstukken.

Hoofdstuk 1 presenteert een algemene inleiding tot sociale media en wetenschappelijke communicatie. Het bespreekt de oorsprong, definities en beschikbaarheid van gegevens (via verschillende aanbieders van altmetrics-gegevens) van sociale mediastatistieken. Het onderzoekt de uitdagingen, beperkingen en mogelijkheden van social media metrics voor onderzoeksevaluatie. Het beschrijft verschillende databronnen voor sociale media, met name Mendeley als een specifieke relevante gegevensbron voor onderzoeksevaluatie. Ten slotte introduceert dit hoofdstuk de hoofddoelstelling en onderzoeksvragen van dit proefschrift.

Hoofdstuk 2 biedt een algemeen overzicht van de aanwezigheid en dekking van publicaties die op sociale mediaplatforms en de verdeling van social media metrics over wetenschappelijke velden, publicatiejaren en documenttypen. Dit hoofdstuk geeft inzicht in de mate waarin wetenschappelijke publicaties worden behandeld op sociale mediaplatformen, de hoeveelheid aandacht die zij ontvangen op sociale media, en disciplinaire verschillen in hun ontvangst op sociale media. Dit hoofdstuk beschrijft ook de relatie tussen social media metrics en citatie-indicatoren. Rekening houdend met alle resultaten gepresenteerd in hoofdstuk 2, ondersteunen de hoge dekking van publicaties, dichtheid (gemiddeld aantal lezers per publicatie) en correlatie van het aantal lezers op Mendeley met de aantallen citaties de conclusie dat indicatoren van het aantal lezers een meer wetenschappelijke impact weerspiegelen, terwijl andere statistieken, zoals van Twitter, Facebook en Wikipedia, de impact op sociale media weerspiegelen.

Hoofdstuk 3 biedt een grondige analyse van de belangrijkste uitdagingen op het gebied van gegevenskwaliteit en problemen met betrekking tot sociale mediadata die worden

aangeboden door de belangrijkste gegevensaanbieders van altmetrics. In dit hoofdstuk wordt besproken hoe de methoden voor gegevensverzameling en rapportage van deze aanbieders van altmetrics zowel de technische als de conceptuele factoren beïnvloeden. Uit de bevindingen blijkt dat de grote aanbieders van altmetrics aanzienlijk verschillende uitkomsten geven van dezelfde sociale mediastatistieken op dezelfde groep DOI's op hetzelfde tijdstip. De belangrijkste uitdagingen met betrekking tot de gegevenskwaliteit van social media metrics worden bepaald door de volgende methodologische keuzes die door de verscheidene gegevensaanbieders van altmetrics zijn genomen: keuzes over gegevensverzameling, keuzes over aggregatie van gegevens en rapportagekeuzes, en het bijwerken van keuzes. Op basis van de resultaten van dit hoofdstuk worden enkele aanbevelingen gegeven. Gegevensaanbieders van altmetrics zouden de transparantie rond de methodologische keuzes in gegevensverzameling, aggregatie en berekening van statistieken moeten vergroten. Gebruikers van altmetrics-gegevens, onderzoekers en gegevensverzamelaars moeten zich bewust zijn van de onbedoelde effecten die deze methodologische keuzes kunnen hebben bij het valide gebruik en de toepassing van social media metrics. Inzicht in de manier waarop methodologische en technische keuzes de analytische betrouwbaarheid en validiteit van social media metrics kunnen beïnvloeden, is een cruciaal element in de toekomstige ontwikkeling van onderzoek naar wetenschap met behulp van sociale media.

Hoofdstuk 4 beschrijft de verschillen tussen wetenschappelijke disciplines in de relatie tussen het aantal lezers op Mendeley en aantallen citaties met bepaalde bibliografische kenmerken van specifieke documenten in een dataset van 1,3 miljoen publicaties in het Web of Science. De associatie tussen het aantal lezers op Mendeley, aantallen citaties en documentkenmerken (dat wil zeggen documenttypen, aantal pagina's, lengte van titels, lengte van referentielijsten, aantal auteurs, instituten en landen) is onderzocht met behulp van Ordinary Least Square (OLS) -regressieanalyse. Dit laat zien dat documenttypen zoals redactionele materialen, brieven, nieuwsitems, boekbesprekingen en meeting abstracts een veel hogere dekking hebben in Mendeley en een veel hogere dichtheid voor lezers dan citaties. Publicaties met relatief hogere Mendeley-lezersaantallen zijn ook gerelateerd aan dezelfde bibliografische kenmerken als die zijn waargenomen voor publicaties met relatief hogere citatietellingen. Het hoofdstuk laat zien welke documentgerelateerde verschillen er zijn in de relatie tussen het aantal lezers op Mendeley en citaties. Deze informatie is nuttig voor de toekomstige constructie van geschikte en betekenisvolle indicatoren op basis van het aantal lezers op Mendeley.

Hoofdstuk 5 presenteert de resultaten van een grootschalige analyse van de spreiding en aanwezigheid van Mendeley-leesscores in de loop van de tijd en tussen disciplines van 9,1 miljoen publicaties in Web of Science in de periode 2004-2013. De resultaten van dit hoofdstuk tonen aan dat lezersaantallen van Mendeley inderdaad effectiever zijn (in termen van precision / recall-waarden) dan op tijdschriftgebaseerde indicatoren bij het filteren van zeer geciteerde publicaties over alle wetenschapsgebieden en publicatiejaren. Geconcludeerd wordt dat lezersindicatoren van Mendeley een effectiever hulpmiddel zijn om veelgeciteerde publicaties te filteren dan op tijdschriftgebaseerde citatie-impactindicatoren. Deze conclusie opent de deur om het lezerspubliek van Mendeley op te nemen als een geschikte en bruikbare indicator voor de voorspelling van toekomstige citaten.

Hoofdstuk 6 richt zich op de verschillende gebruikersgroepen in Mendeley en hun thematische oriëntaties. Een dataset van 1,1 miljoen Web of Science-publicaties uit het jaar 2012 wordt geanalyseerd. De disciplinaire verschillen in de lees- en opslaanpatronen van verscheidene Mendeley-gebruikersgroepen worden weergegeven met behulp van VOSviewer-kaarten. De interessegebieden van verschillende gebruikersgroepen in Mendeley worden geanalyseerd. De resultaten van dit hoofdstuk geven aan dat verscheidene gebruikersgroepen relatief meer aandacht hebben voor publicaties met betrekking tot hun functie en het doel waarvoor ze Mendeley gebruiken. De resultaten wijzen op het idee dat publicaties die zijn opgeslagen door verscheidene gebruikersgroepen kunnen worden gerelateerd aan verschillende contexten van gebruik, zoals onderwijs, (zelf)training, onderzoek, en praktisch en toegepast gebruik. Deze resultaten suggereren dat de analyse van het leesgedrag van verschillende Mendeley-gebruikersgroepen kan worden gebruikt om hun wetenschappelijke, beroepsmatige en onderwijsinteresses te identificeren. De resultaten in hoofdstuk 6 benadrukken ook de mogelijke functie van lezersindicatoren om het gebruik van wetenschappelijke documenten door een groot aantal doelgroepen vast te leggen.

Tot slot bevat **hoofdstuk 7** de discussie en conclusies van de belangrijkste resultaten van dit proefschrift. Het presenteert de samenvatting van de bevindingen en de implicaties van de verkregen resultaten voor onderzoeksevaluaties, samen met enkele perspectieven voor verder onderzoek. De verscheidene resultaten die in dit proefschrift worden gepresenteerd, tonen duidelijk aan dat analyse van het lezerspubliek van Mendeley de sociale mediabron is met de grootste bruikbaarheid en toegevoegde waarde voor onderzoeksevaluatie. Dit is gerechtvaardigd op basis van de grote dekking, dichtheid, correlatie, documentkenmerken en conceptuele gelijkenis van Mendeley-lezers met citatie-indicatoren. Deze toegevoegde waarde van het Mendeley-lezerspubliek voor onderzoeksevaluatie wordt bijzonder duidelijk als deze wordt vergeleken met andere sociale mediastatistieken (bijvoorbeeld Twitter-, Facebook- of Wikipedia-aantallen). De laatste hebben een meer marginale dekking en dichtheid, veel lagere correlatie met aantallen citaties en meer fundamentele conceptuele verschillen met het gebruik van citaties. Enkele mogelijke richtingen voor verder onderzoek op basis van dit onderzoek worden gepresenteerd en besproken, met name in de richting van de ontwikkeling van geavanceerdere lezersindicatoren (bijvoorbeeld door gedetailleerdere informatie over de interacties van Mendeley-gebruikers met wetenschappelijke output op te nemen), en van de ontwikkeling van theorie om het gedrag van online lezers beter te begrijpen.

Appendix 1 (Supplementary information to Chapter 3)**:**

Texts 1-5. Extracts of methodological descriptions from Social Media platforms and Altmetric aggregators.

Text 1

"The Twitter Search API is part of Twitter's REST API. It allows queries against the indices of recent or popular Tweets and behaves similarly to, but not exactly like the Search feature available in Twitter mobile or web clients, such as Twitter.com search. The Twitter Search API searches against a sampling of recent Tweets published in the past 7 days. Before getting involved, it's important to know that the Search API is focused on relevance and not completeness. This means that some Tweets and users may be missing from search results. If you want to match for completeness you should consider using a Streaming API instead[1]."

Text 2

"Altmetric.com tracks Twitter attention in real-time via an API. We collect tweets, retweets, and quoted tweets that contain a direct link to a scholarly output from a publisher we are tracking[2]. It reports public tweets, retweets but not favorites or likes, that link directly to research outputs (protected tweets and any tweets prior to 2011 are not tracked)".

"We [Plum Analytics] license twitter data in PlumX directly through Twitter/GNIP. We have a filtered view of all tweets based upon the domain names of the links in the tweets. Our historic twitter data begins on January 1, 2011. We accommodate URL shorteners and have match and merge technology for combining tweets from multiple, separate URLs into a single view for a given artifact [object]. However, if the original artifact is published at a domain that we do not yet track, once identified and added by the Plum Analytics team, twitter mentions for that domain will only begin to be counted from the time the new domain is added[3]".

"We [CrossRef ED] submit a set of filter rules to the Gnip Power Track service. This list is made up of DOI prefixes from the doi-prefix-list Artifact, DOI resolver domains, all domains in the domain-list Artifact (e.g., journals.iucr.org). The rules sent to Gnip Power Tack are manually updated. We aim to keep them in sync with the domain-list Artifact, but they may lag slightly. The Agent monitors all data sent back from the Power Track stream. This includes tweets that contain a DOI prefix, a hyperlinked DOI, Landing Page URL, or a link-shortened link to a DOI or Landing Page URL. The Gnip service automatically follows and extracts URLs from link-shortening services like bit.ly before the data is sent to us. This gives the Twitter source an advantage, as it removes opaque link-shortened links that we otherwise could not match. We then attempt to match all links to Registered Content Items. Publisher sites may block the Event Data Bot collecting Landing Pages[4]".

---

1 https://developer.twitter.com/en/docs/tweets/rules-and-filtering/guides/how-to-build-a-query

2 https://help.altmetric.com/support/solutions/articles/6000157183-how-does-altmetric-track-twitter-

3 https://plumanalytics.com/niso-altmetrics-working-group-on-data-quality/

4 www.eventdata.crossref.org/guide/sources-in-depth/

"Lagotto searches the Twitter Search API by DOI and URL, e.g., The Search API will find shortened URLs with this query. The rate-limits for application-only authentication and search are 450 requests per 15 min or 1,800 requests per hour. Depending on the number of articles we might have to adjust how often we contact Twitter, the default settings are every 12 hours the first 7 days after publication, then daily for the first month, and then weekly[5]".

Text 3

Plum Analytics:

"Online events about different versions of the same artifact [object](Publisher + Green Open Access + Preprint + Aggregated versions + A&I) are collected and aggregated based on algorithms that examine matching identifiers (such as DOI, ISBN, or URI) across versions. Usage, Capture, Social Media, and Mention metrics counts are summed across all versions of each artifacts. Our match and merge algorithms for combining and aggregating metrics from all the different online locations where it is published depend upon a knowledge base of how to cross-walk different identifiers (like going from a DOI to a PubMed ID). If there are errors in this crosswalk data, it is possible to "over-merge" a record. Similarly, if there is not enough data to automatically merge two preprints from two different services together, they may also need to be manually identified and merged by the PlumX staff[6]".

Text 4

"When a user deletes its account and all their documents, the readership of that document doesn't change, until the batch clustering process is re-run and the new number of metadata records is generated. The same applies when a user deletes a record from its library. In summary, the count of records can increase nearly instantaneously, but only decreases periodically[7]".

Text 5

"These identifiers also help us [altmetric.com] to recognize different versions of the same research output. For example, a journal article might be originally made available on a publisher platform and given a DOI, and then later hosted on PubMed or an institutional repository and given another unique identifier there. Our system cross-checks these to match them together, ensuring that the details page always displays a collated record of attention for all versions of the research item[8]".

[5] http://www.lagotto.io/docs/twitter_search/

6 https://plumanalytics.com/niso-altmetrics-working-group-on-data-quality/

7 www.niso.org/apps/group_public/view_comment.php?comment_id=632;
www.niso.org/apps/group_public/view_comment.php?comment_id=610

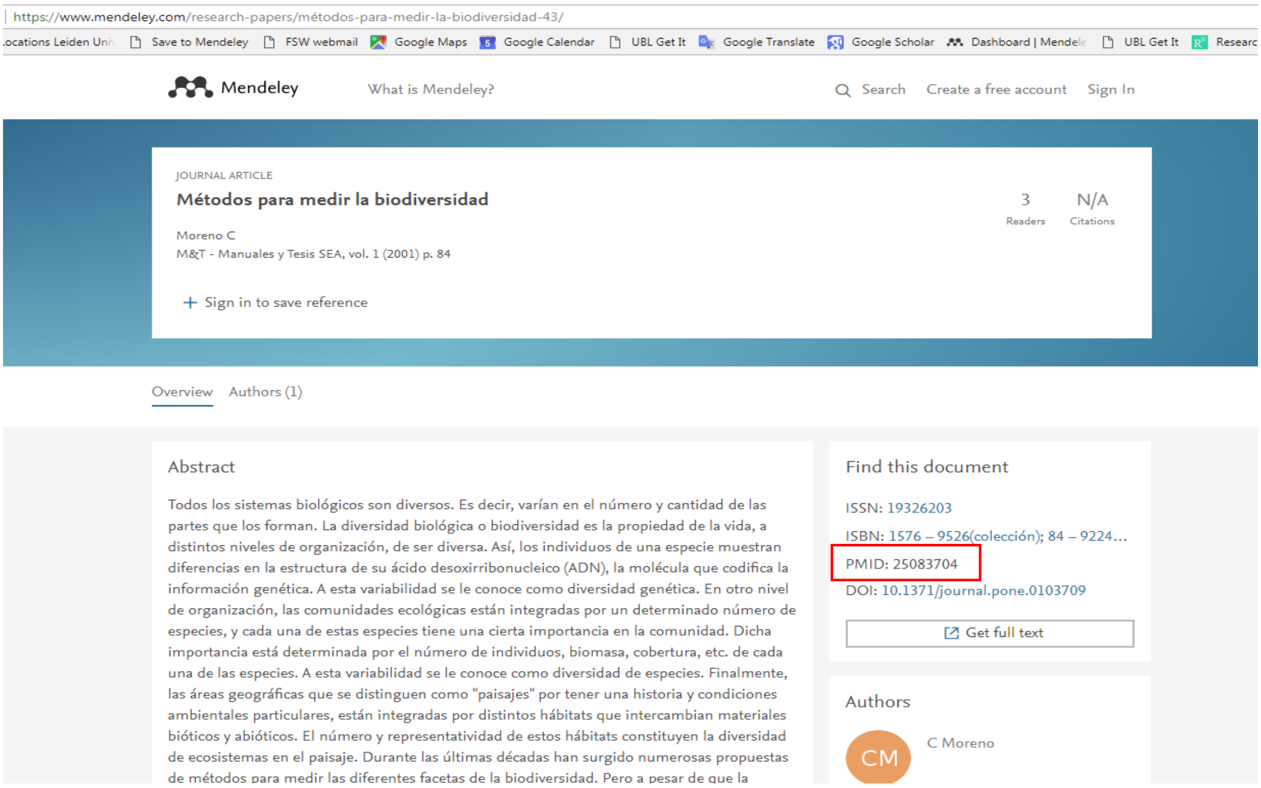
[8] https://www.altmetric.com/about-our-data/how-it-works/

Figures 1-2. Examples of records in Mendeley with incorrect PMIDs

Figure 1. *Examples of records in Mendeley with incorrect PMDI 25083704:*

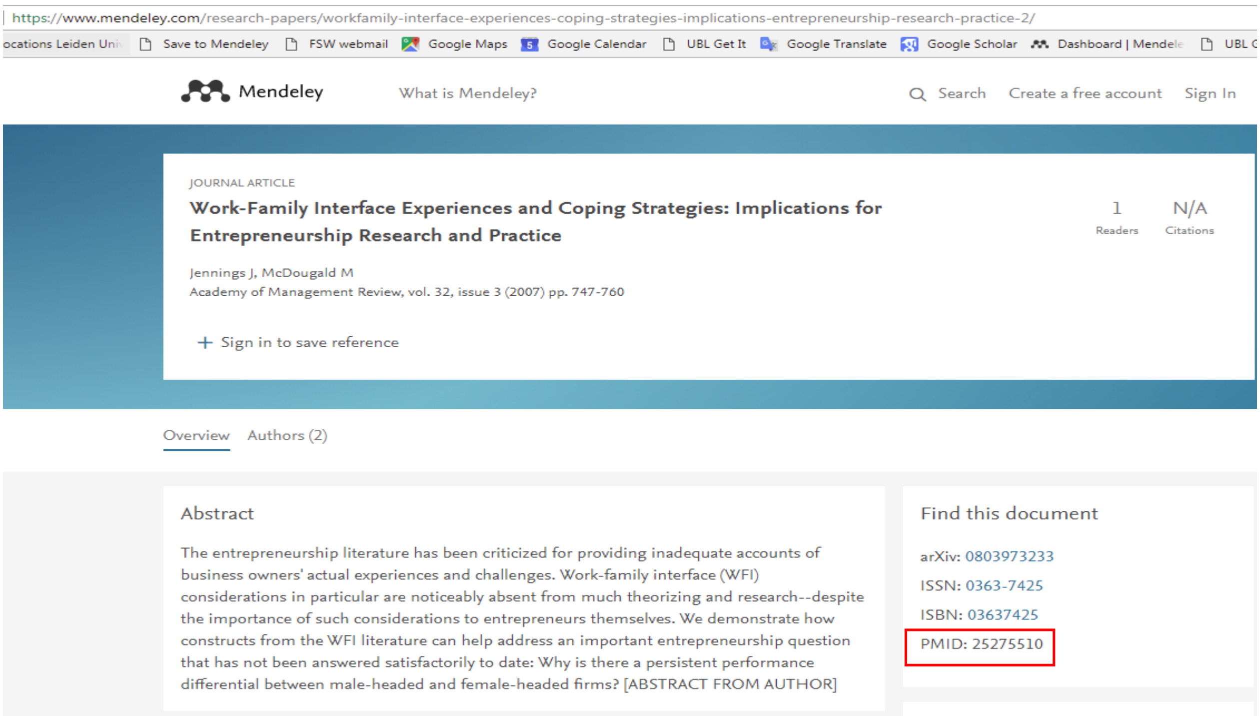
https://www.mendeley.com/research-papers/m%C3%A9todos-para-medir-la-biodiversidad-23/

https://www.mendeley.com/research-papers/systematic-distribution-orihelia-anticlava-molin-1858-nematoda-onchocercidae-dasypodids-south-americ/



https://www.mendeley.com/research-papers/m%C3%A9todos-para-medir-la-biodiversidad-43/

Figure 2. An example of a record in Mendeley with incorrect PMDI=25275510:



https://www.mendeley.com/research-papers/workfamily-interface-experiences-coping-strategies-implications-entrepreneurship-research-practice-2/

Text 6. Excerpt of the JSON file from CrossRef ED recording 315 Wikipedia mentions for an object with DOI: 10.1371.journal.pone.0105090):

{"status":"ok","message-type":"event-list","message":{"next-cursor":"0d2917ad-89d6-4bd5-a57b-36af574e8510","total-results":315,"items-per-page":1000,"events":[{"license":"https:\/\/creativecommons.org\/publicdomain\/zero\/1.0\/","obj_id":"https:\/\/doi.org\/10.1371\/journal.pone.0105090","source_token":"36c35e23-8757-4a9d-aacf-345e9b7eb50d","occurred_at":"2017-04-18T14:49:16Z","subj_id":"https:\/\/mk.wikipedia.org\/w\/index.php?title=%D0%9C%D0%B0%D0%BA%D0%B5%D0%B4%D0%BE%D0%BD%D1%86%D0%B8&oldid=3586676","id":"82ed5180-b2ee-486a-b495-aa2883c1b1be","evidence_record":"https:\/\/evidence.eventdata.crossref.org\/evidence\/20170418-wikipedia-4a99bd66-e33f-49c5-a52a-0b3690c3e5a0","terms":"https:\/\/doi.org\/10.13003\/CED-terms-of-use","action":"add","subj":{"pid":"https:\/\/mk.wikipedia.org\/w\/index.php?title=%D0%9C%D0%B0%D0%BA%D0%B5%D0%B4%D0%BE%D0%BD%D1%86%D0%B8&oldid=3586676","url":"https:\/\/mk.wikipedia.org\/wiki\/%D0%9C%D0%B0%D0%BA%D0%B5%D0%B4%D0%BE%D0%BD%D1%86%D0%B8","title":"\u041c\u0430\u043a\u0435\u0434\u043e\u043d\u0446\u0438","api-url":"https:\/\/mk.wikipedia.org\/api\/rest_v1\/page\/html\/%D0%9C%D0%B0%D0%BA%D0%B5%D0%B4%D0%BE%D0%BD%D1%86%D0%B8\/3586676"},"source_id":"wikipedia","obj":{"pid":"https:\/\/doi.org\/10.1371\/journal.pone.0105090","url":"https:\/\/doi.org\/10.1371\/journal.pone.0105090"},"timestamp":"2017-04-18T14:50:20Z","relation_type_id":"references"}, ….

Appendix 2 (Supplementary information to Chapter 4):

Table A1. Descriptive table of the independent and the dependent variables across articles and reviews disciplines.

Leiden Ranking Discipline		PG	NR	TI	AU	IN	CU
All disciplines	Density	10.06	39.60	98.75	5.42	2.21	1.34
N=1,197,162	*Std. dev.*	7.60	31.22	35.05	37.70	3.75	0.98
Biomedical & health sciences	Density	8.75	41.30	104.51	5.98	2.45	1.33
N=502,610	*Std. dev.*	*6.81*	*33.03*	*36.51*	*5.11*	*2.50*	*0.88*
Life & earth sciences	Density	10.28	45.55	106.43	4.96	2.28	1.41
N=242,117	*Std. dev.*	*6.56*	*31.20*	*33.88*	*4.27*	*1.78*	*0.83*
Mathematics & computer science	Density	13.06	28.14	81.91	3.21	1.82	1.31
N=131,220	*Std. dev.*	*8.40*	*19.32*	*27.64*	*2.81*	*1.16*	*0.61*
Natural sciences & engineering	Density	8.94	36.08	97.53	6.38	2.12	1.35
N=404,457	*Std. dev.*	*6.58*	*29.13*	*33.06*	*64.55*	*5.70*	*1.25*
Social sciences & humanities	Density	15.02	46.32	87.15	2.77	1.83	1.26
N=125,792	*Std. dev.*	*8.18*	*30.16*	*32.10*	*2.45*	*1.38*	*0.64*

PG= Number of Page; NR= Number of Reference, TI=Title Length, AU=Number of Author, IN=Number of Institute, CU=Number of Country.

Table A2. Spearmann correlation analysis among the independent and the dependent variables (C and MR) across articles and reviews by disciplines.

Biomedical & health sciences (n= 502,610)

	PG	NR	TI	AU	IN	CU	C	MR
PG	1.000	0.652	0.092	0.052	0.105	0.132	0.318	0.338
NR		1.000	-0.008	-0.031	0.064	0.115	0.381	0.392
TI			1.000	0.253	0.107	0.046	0.052	-0.033
AU				1.000	0.452	0.243	0.186	0.015
IN					1.000	0.511	0.156	0.122
CU						1.000	0.165	0.164
C							1.000	0.558
MR								1.000

Life & earth sciences (n=242,117)

	PG	NR	TI	AU	IN	CU	C	MR
PG	1.000	.560	.069	-.041	.076	.106	.149	.166
NR		1.000	.067	.031	.107	.133	.345	.383
TI			1.000	.156	.064	.019	.008	-.065
AU				1.000	.508	.272	.196	.070
IN					1.000	.579	.142	.132
CU						1.000	.148	.163
C							1.000	.585
MR								1.000

Natural sciences & engineering (n=404,457)

	PG	NR	TI	AU	IN	CU	C	MR
PG	1.000	.424	.020	-.127	.051	.087	.081	.095
NR		1.000	.101	.061	.089	.121	.429	.341
TI			1.000	.157	.006	-.043	.073	0.001
AU				1.000	.434	.246	.185	.132
IN					1.000	.597	.097	.087
CU						1.000	.126	.110
C							1.000	.581
MR								1.000

Mathematics & computer sciences (n= 131,220)

	PG	NR	TI	AU	IN	CU	C	MR
PG	1.000	.430	-.071	-.168	.054	.095	.022	.034
NR		1.000	.122	.122	.136	.111	.336	.424
TI			1.000	.184	.040	-.027	.119	.080
AU				1.000	.454	.251	.193	.242
IN					1.000	.601	.119	.123
CU						1.000	.104	.091
C							1.000	.452
MR								1.000

Social sciences & humanities (n=125,795)

	PG	NR	TI	AU	IN	CU	C	MR
PG	1.000	.412	-.041	-.264	-.114	-.017	-.103	-.081
NR		1.000	.161**	.075	.079	.073	.286	.322
TI			1.000	.273	.146	.043	.157	.138
AU				1.000	.635	.326	.396	.339
IN					1.000	.570	.289	.247
CU						1.000	.178	.175
C							1.000	.638
MR								1.000

PG= Number of Page; NR= Number of Reference, TI=Title Length, AU=Number of Author, IN=Number of Institute, CU=Number of Country, C= Citation, MR= Mendeley Readership

Appendix 3 (Supplementary information to Chapter 5):

Table 1. General distributions of MRS, MCS & JCS over WoS publications across LR fields and publication years 2004-2013.

(Number of publications (P); Total Readership Score (TRS); Mean Readership Score (MRS); Total Citation Score (TCS); Mean Citation Score (MCS); Coverage (n.pub) in Mendeley (Cov))

LR Main fields	Pub year	P	%	TRS	MRS	TCS	MCS	JCS	Cov
Biomedical & health sciences	**2004**	232,480	37.37	2,884,544	12.41	8,409,470	36.17	35.99	87.48
Life & earth sciences		88,656	14.25	1,815,332	20.48	3,021,126	34.08	33.86	91.38
Mathematics & computer science		53,724	8.64	391,561	7.29	786,012	14.63	14.35	75.79
Natural sciences &engineering		214,396	34.47	1,619,620	7.55	5,305,063	24.74	24.69	81.40
Social sciences & humanities		32,784	5.27	701,348	21.39	769,230	23.46	23.40	91.25
Biomedical & health sciences	**2005**	262,863	36.86	3,543,882	13.48	8,929,096	33.97	33.84	88.72
Life & earth sciences		107,303	15.05	2,303,379	21.47	3,370,412	31.41	31.31	92.07
Mathematics & computer science		65,236	9.15	467,938	7.17	834,529	12.79	12.79	75.90
Natural sciences &engineering		236,252	33.13	1,879,802	7.96	5,564,041	23.55	23.58	82.88
Social sciences & humanities		41,494	5.82	900,913	21.71	876,384	21.12	21.03	91.31
Biomedical & health sciences	**2006**	296,670	36.19	4,115,600	13.87	9,038,703	30.47	30.43	89.66
Life & earth sciences		127,610	15.57	2,688,317	21.07	3,443,035	26.98	26.87	92.62
Mathematics & computer science		74,272	9.06	564,160	7.60	880,429	11.85	11.62	75.66
Natural sciences &engineering		271,343	33.10	2,146,430	7.91	5,589,198	20.60	20.60	82.76
Social sciences & humanities		49,871	6.08	1,052,977	21.11	908,119	18.21	18.13	91.18
Biomedical & health sciences	**2007**	327,757	36.00	4,680,701	14.28	8,871,383	27.07	26.94	90.38
Life & earth sciences		146,240	16.06	3,045,174	20.82	3,525,421	24.11	23.99	93.14
Mathematics & computer science		85,404	9.38	680,572	7.97	912,715	10.69	10.49	76.76
Natural sciences &engineering		291,797	32.05	2,378,578	8.15	5,422,527	18.58	18.64	82.47
Social sciences & humanities		59,299	6.51	1,261,111	21.27	952,984	16.07	15.82	91.39

LR Main fields	Pub year	P	%	TRS	MRS	TCS	MCS	JCS	Cov
Biomedical & health sciences	**2008**	359,473	35.70	5,311,342	14.78	8,333,098	23.18	23.14	91.70
Life & earth sciences		164,826	16.37	3,451,344	20.94	3,414,915	20.72	20.68	94.03
Mathematics & computer science		94,443	9.38	764,848	8.10	868,684	9.20	9.20	77.98
Natural sciences &engineering		308,194	30.61	2,672,742	8.67	5,300,704	17.20	17.22	84.33
Social sciences & humanities		79,884	7.93	1,578,239	19.76	989,360	12.38	12.32	92.59
Biomedical & health sciences	**2009**	394,055	35.39	5,998,591	15.22	7,754,064	19.68	19.59	92.73
Life & earth sciences		181,092	16.26	3,674,016	20.29	3,151,425	17.40	17.38	94.60
Mathematics & computer science		106,493	9.56	877,863	8.24	845,048	7.94	7.89	78.81
Natural sciences &engineering		338,831	30.43	3,057,071	9.02	5,124,469	15.12	15.14	85.56
Social sciences & humanities		92,996	8.35	1,819,415	19.56	945,439	10.17	10.06	93.95
Biomedical & health sciences	**2010**	426,760	35.99	6,441,730	15.09	6,669,619	15.63	15.56	92.76
Life & earth sciences		196,995	16.61	3,827,502	19.43	2,739,005	13.90	13.88	94.25
Mathematics & computer science		108,352	9.14	889,800	8.21	694,745	6.41	6.40	78.25
Natural sciences &engineering		352,548	29.73	3,282,464	9.31	4,478,561	12.70	12.71	85.48
Social sciences & humanities		101,219	8.54	1,915,263	18.92	799,615	7.90	7.84	93.07
Biomedical & health sciences	**2011**	458,101	35.29	6,262,656	13.67	5,255,365	11.47	11.42	92.29
Life & earth Sciences		219,838	16.94	3,697,421	16.82	2,245,003	10.21	10.17	93.45
Mathematics & computer science		118,363	9.12	885,505	7.48	578,200	4.88	4.87	77.25
Natural sciences &engineering		387,280	29.84	3,275,521	8.46	3,758,759	9.71	9.71	84.48
Social sciences & humanities		114,430	8.82	1,860,287	16.26	628,106	5.49	5.48	92.87
Biomedical & health sciences	**2012**	495,779	35.65	5,392,213	10.88	3,702,869	7.47	7.45	89.16
Life & earth sciences		239,785	17.24	3,222,091	13.44	1,573,666	6.56	6.52	91.23
Mathematics & computer science		128,429	9.23	804,163	6.26	415,987	3.24	3.21	74.57
Natural sciences &engineering		404,991	29.12	2,866,338	7.08	2,751,787	6.79	6.80	82.52

LR Main fields	Pub year	P	%	TRS	MRS	TCS	MCS	JCS	Cov
Social sciences & humanities		121,888	8.76	1,618,761	13.28	426,640	3.50	3.49	91.45
Biomedical & health sciences	**2013**	86,899	35.47	837,117	9.63	474,055	5.46	3.79	90.85
Life & earth sciences		39,828	16.26	464,543	11.66	184,160	4.62	3.22	92.38
Mathematics & computer science		24,647	10.06	144,169	5.85	60,686	2.46	1.76	78.58
Natural sciences &engineering		73,350	29.94	463,308	6.32	360,998	4.92	3.52	84.05
Social sciences & humanities		20,277	8.28	248,331	12.25	50,328	2.48	1.73	93.78

Figure 1. Distribution of MRS vs. MCS by LR fields across publication years (2004-2013).

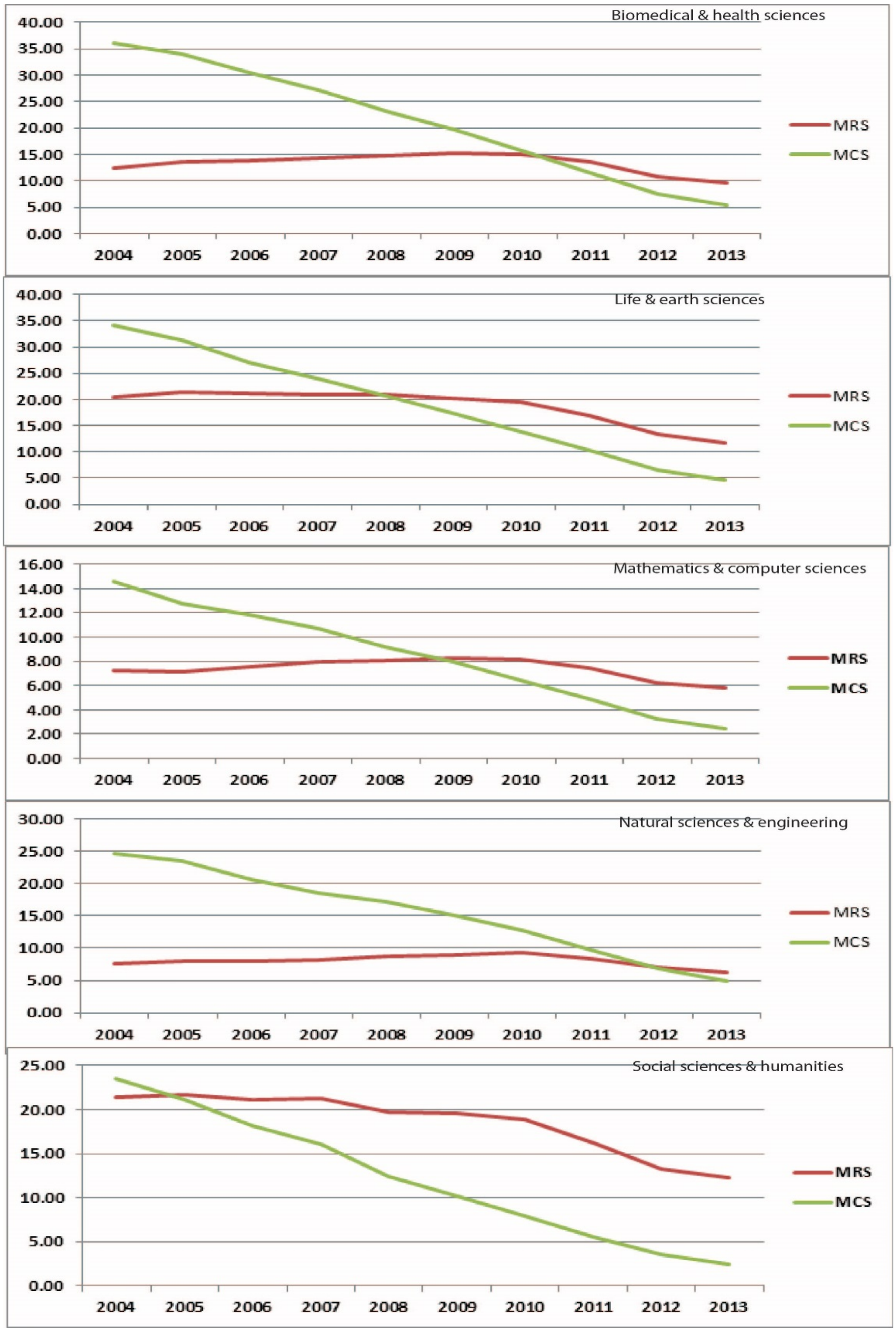


Table 2. Distributions of MRS vs. MCS over WoS publications across 250 WoS Subject categories sorted alphabetically.

(Number of publications (P); Total Readership Score (TRS); Mean Readership Score (MRS); Total Citation Score (TCS); Mean Citation Score (MCS))

WoS subject category	P	TCS	MCS	TRS	MRS
acoustics	15,958	126,301	7.91	123,703	7.75
automation & control systems	22,796	177,262	7.78	116,191	5.10
agriculture, dairy & animal science	25,986	209,166	8.05	170,422	6.56
agricultural economics & policy	1,596	9,575	6.00	17,689	11.08
agriculture, multidisciplinary	17,902	182,389	10.19	155,602	8.69
engineering, aerospace	12,727	46,568	3.66	61,990	4.87
agronomy	26,407	251,804	9.54	252,990	9.58
allergy	8,518	148,806	17.47	64,025	7.52
anatomy & morphology	9,422	91,653	9.73	98,793	10.49
andrology	2,760	29,254	10.60	13,070	4.74
anesthesiology	22,293	255,155	11.45	203,621	9.13
biodiversity conservation	9,605	139,632	14.54	361,867	37.67
anthropology	9,391	79,028	8.42	183,037	19.49
archaeology	5,448	26,856	4.93	71,016	13.03
architecture	1,864	998	0.54	5,123	2.75
area studies	7,461	20,036	2.69	42,162	5.65
art	2,423	2,097	0.87	8,859	3.66
humanities, multidisciplinary	7,423	5,348	0.72	19,040	2.57
astronomy & astrophysics	104,601	1,809,038	17.29	529,229	5.06
psychology, biological	3,020	42,798	14.17	55,227	18.28
behavioral sciences	19,246	331,287	17.21	495,083	25.72
biochemical research methods	62,039	1,230,789	19.84	1,184,037	19.09
biochemistry & molecular biology	271,026	6,458,150	23.83	4,408,796	16.27
biology	33,529	563,130	16.80	808,356	24.11
biophysics	52,205	926,360	17.74	722,858	13.85
biotechnology & applied microbiology	90,278	1,631,317	18.07	1,674,400	18.55
plant sciences	91,807	1,548,338	16.87	1,544,551	16.82
business	19,730	202,516	10.26	616,481	31.25
business, finance	13,767	114,054	8.28	228,296	16.58
oncology	168,747	4,007,839	23.75	1,577,205	9.35
cardiac & cardiovascular systems	96,242	1,892,030	19.66	749,475	7.79
cell biology	112,976	3,420,859	30.28	2,544,510	22.52
thermodynamics	21,104	193,908	9.19	137,225	6.50
chemistry, applied	37,025	451,142	12.18	246,759	6.66
chemistry, medicinal	45,827	681,456	14.87	344,433	7.52
chemistry, multidisciplinary	204,339	5,620,720	27.51	2,541,258	12.44
chemistry, analytical	93,824	1,472,917	15.70	728,830	7.77
chemistry, inorganic & nuclear	80,388	1,124,811	13.99	358,053	4.45
chemistry, organic	133,126	2,054,966	15.44	726,230	5.46
chemistry, physical	200,438	3,288,968	16.41	1,660,898	8.29
classics	1,495	997	0.67	1,618	1.08
computer science, artificial intelligence	34,362	262,549	7.64	396,374	11.54
psychology, clinical	26,733	370,534	13.86	440,892	16.49
computer science, cybernetics	5,230	38,951	7.45	64,975	12.42
computer science, hardware & architecture	10,789	39,459	3.66	71,023	6.58

WoS subject category	P	TCS	MCS	TRS	MRS
computer science, information systems	33,393	213,348	6.39	371,466	11.12
communication	10,742	72,042	6.71	199,290	18.55
computer science, interdisciplinary applications	33,189	282,256	8.50	369,071	11.12
computer science, software engineering	26,302	122,366	4.65	275,586	10.48
computer science, theory & methods	25,028	122,827	4.91	189,348	7.57
construction & building technology	13,103	98,707	7.53	122,892	9.38
criminology & penology	8,203	55,969	6.82	76,952	9.38
emergency medicine	16,705	129,055	7.73	126,783	7.59
crystallography	61,996	358,424	5.78	110,510	1.78
dance	287	222	0.77	696	2.43
demography	3,308	21,601	6.53	34,123	10.32
dentistry/oral surgery & medicine	45,189	421,268	9.32	482,177	10.67
dermatology	34,710	367,830	10.60	176,778	5.09
geochemistry & geophysics	58,348	844,120	14.47	685,619	11.75
substance abuse	14,671	195,800	13.35	127,513	8.69
ecology	58,822	975,999	16.59	2,270,219	38.59
economics	69,254	480,111	6.93	788,363	11.38
education & educational research	33,177	176,773	5.33	565,037	17.03
education, scientific disciplines	9,052	56,574	6.25	106,785	11.80
education, special	3,777	26,184	6.93	42,193	11.17
psychology, educational	8,218	86,090	10.48	155,114	18.88
electrochemistry	46,923	764,836	16.30	409,397	8.72
evolutionary biology	18,962	421,024	22.20	745,393	39.31
developmental biology	23,143	613,501	26.51	489,681	21.16
endocrinology & metabolism	96,049	2,014,503	20.97	882,521	9.19
energy & fuels	52,998	772,238	14.57	746,567	14.09
engineering, multidisciplinary	23,249	141,533	6.09	121,291	5.22
engineering, biomedical	36,889	546,637	14.82	518,980	14.07
engineering, environmental	30,863	523,515	16.96	379,234	12.29
engineering, chemical	98,531	1,134,783	11.52	686,942	6.97
engineering, industrial	11,903	98,278	8.26	141,199	11.86
engineering, manufacturing	16,781	120,965	7.21	128,405	7.65
engineering, marine	1,847	5,531	2.99	6,169	3.34
engineering, civil	39,468	316,984	8.03	289,745	7.34
engineering, ocean	2,345	17,378	7.41	16,482	7.03
engineering, petroleum	1,972	7,036	3.57	8,540	4.33
engineering, electrical & electronic	184,091	1,340,863	7.28	1,038,344	5.64
engineering, mechanical	56,984	388,294	6.81	303,794	5.33
entomology	23,161	187,456	8.09	223,791	9.66
environmental sciences	123,365	1,780,522	14.43	1,837,151	14.89
environmental studies	16,921	164,938	9.75	368,727	21.79
ergonomics	2,665	20,172	7.57	50,024	18.77
ethnic studies	2,127	10,799	5.08	21,372	10.05
family studies	5,999	49,958	8.33	69,377	11.56
film, radio, television	1,776	1,902	1.07	5,799	3.26
fisheries	20,625	186,918	9.06	271,668	13.17
folklore	410	239	0.58	594	1.45
food science & technology	70,919	810,317	11.43	597,364	8.42
forestry	20,527	215,829	10.51	321,283	15.65
gastroenterology & hepatology	68,407	1,356,545	19.83	441,786	6.46
genetics & heredity	89,903	2,145,766	23.87	2,030,277	22.58

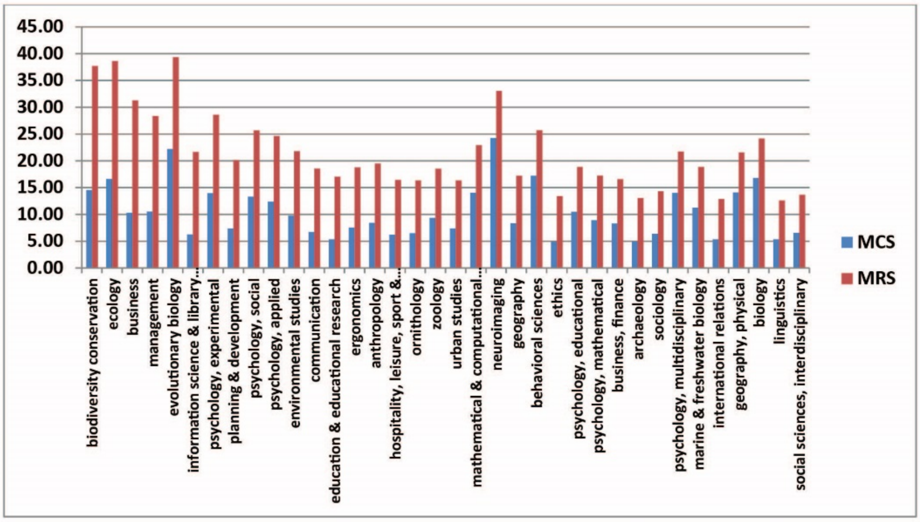
WoS subject category	P	TCS	MCS	TRS	MRS
geography	12,347	103,564	8.39	212,153	17.18
geography, physical	13,556	190,780	14.07	292,374	21.57
geology	10,199	139,376	13.67	135,022	13.24
geosciences, multidisciplinary	79,755	1,031,473	12.93	1,026,675	12.87
geriatrics & gerontology	17,650	236,213	13.38	188,362	10.67
health policy & services	12,927	134,977	10.44	160,386	12.41
hematology	57,589	1,264,167	21.95	538,681	9.35
history	15,223	18,288	1.20	35,631	2.34
history & philosophy of science	6,977	23,169	3.32	61,444	8.81
history of social sciences	2,733	6,501	2.38	13,261	4.85
horticulture	7,031	64,330	9.15	61,593	8.76
psychology, developmental	19,575	297,081	15.18	389,249	19.88
public, environmental & occupational health	89,499	1,139,524	12.73	930,221	10.39
immunology	89,847	2,037,686	22.68	1,223,376	13.62
industrial relations & labor	2,347	10,709	4.56	23,699	10.10
infectious diseases	41,119	590,512	14.36	384,097	9.34
psychology, applied	11,877	147,035	12.38	292,727	24.65
information science & library science	11,539	72,320	6.27	249,854	21.65
instruments & instrumentation	39,208	286,205	7.30	201,364	5.14
international relations	9,618	51,663	5.37	123,945	12.89
law	7,034	24,907	3.54	43,509	6.19
medicine, legal	9,377	71,661	7.64	86,515	9.23
asian studies	2,472	1,964	0.79	5,894	2.38
linguistics	11,043	59,010	5.34	139,024	12.59
limnology	5,280	49,249	9.33	62,651	11.87
language & linguistics theory	7,657	20,083	2.62	56,831	7.42
literary reviews	649	314	0.48	1,178	1.82
literature	8,145	4,951	0.61	12,489	1.53
management	29,117	306,676	10.53	825,179	28.34
literature, african, australian, canadian	265	207	0.78	282	1.06
operations research & management science	29,766	254,818	8.56	317,109	10.65
literature, american	748	856	1.15	1,603	2.14
literature, british isles	447	159	0.36	318	0.71
literature, german, dutch, scandinavian	609	177	0.29	293	0.48
marine & freshwater biology	42,275	475,447	11.25	797,562	18.87
materials science, paper & wood	2,533	18,937	7.48	10,815	4.27
materials science, ceramics	28,095	229,266	8.16	130,504	4.65
materials science, multidisciplinary	216,228	2,880,321	13.32	1,755,873	8.12
mathematics, applied	104,978	598,838	5.70	251,266	2.39
mathematics, interdisciplinary applications	23,727	161,591	6.81	113,987	4.80
mathematics	110,828	408,500	3.69	124,022	1.12
social sciences, mathematical methods	4,970	37,600	7.57	48,480	9.75
medical informatics	6,727	60,394	8.98	98,888	14.70
mechanics	61,643	513,163	8.32	356,883	5.79
medical laboratory technology	17,444	206,847	11.86	99,436	5.70
medicine, general & internal	77,935	1,240,741	15.92	779,214	10.00
metallurgy & metallurgical engineering	47,126	358,035	7.60	201,018	4.27

WoS subject category	P	TCS	MCS	TRS	MRS
medicine, research & experimental	57,193	1,068,768	18.69	594,358	10.39
literature, romance	1,383	439	0.32	1,091	0.79
literature, slavic	585	120	0.21	404	0.69
materials science, biomaterials	19,573	367,388	18.77	246,107	12.57
materials science, characterization & testing	7,399	36,573	4.94	28,846	3.90
materials science, coatings & films	22,327	240,425	10.77	145,859	6.53
materials science, composites	14,747	162,490	11.02	120,349	8.16
materials science, textiles	4,752	31,112	6.55	18,242	3.84
medieval & renaissance studies	1,647	1,313	0.80	2,481	1.51
meteorology & atmospheric sciences	52,494	771,111	14.69	560,821	10.68
microbiology	99,535	1,921,385	19.30	1,517,302	15.24
microscopy	6,636	61,684	9.30	65,449	9.86
mineralogy	9,278	95,229	10.26	55,186	5.95
multidisciplinary sciences	170,838	5,969,776	34.94	6,131,068	35.89
music	3,726	5,485	1.47	16,909	4.54
mycology	9,524	106,447	11.18	99,185	10.41
clinical neurology	90,586	1,325,551	14.63	930,957	10.28
neurosciences	177,560	4,158,008	23.42	4,236,027	23.86
nuclear science & technology	33,024	173,548	5.26	99,774	3.02
nursing	27,492	155,110	5.64	214,611	7.81
nutrition & dietetics	40,496	601,944	14.86	432,716	10.69
obstetrics & gynecology	57,277	616,442	10.76	320,166	5.59
oceanography	24,548	306,458	12.48	396,403	16.15
remote sensing	8,080	87,055	10.77	111,418	13.79
ophthalmology	52,435	628,707	11.99	374,450	7.14
optics	94,690	729,840	7.71	598,003	6.32
ornithology	4,672	30,260	6.48	76,389	16.35
orthopedics	40,111	409,275	10.20	373,055	9.30
otorhinolaryngology	27,564	177,351	6.43	114,687	4.16
paleontology	10,196	80,815	7.93	107,717	10.56
parasitology	22,483	295,381	13.14	276,944	12.32
pathology	37,698	488,010	12.95	227,479	6.03
pediatrics	68,218	731,988	10.73	472,234	6.92
pharmacology & pharmacy	148,114	2,467,706	16.66	1,314,788	8.88
philosophy	13,546	24,886	1.84	70,857	5.23
physics, applied	231,661	2,695,659	11.64	1,656,579	7.15
imaging science & photographic technology	7,127	76,555	10.74	87,734	12.31
physics, fluids & plasmas	42,422	419,850	9.90	259,261	6.11
physics, atomic, molecular & chemical	83,180	1,129,941	13.58	600,712	7.22
physics, multidisciplinary	143,341	2,375,872	16.57	1,345,386	9.39
physics, condensed matter	151,594	1,966,429	12.97	1,189,753	7.85
physiology	44,301	880,018	19.86	574,896	12.98
physics, nuclear	27,761	297,210	10.71	69,443	2.50
physics, particles & fields	46,948	648,900	13.82	124,780	2.66
planning & development	9,676	71,537	7.39	194,541	20.10
physics, mathematical	45,592	419,480	9.20	241,707	5.30
poetry	87	15	0.17	22	0.25
political science	21,819	127,096	5.82	266,110	12.20
polymer science	105,037	1,500,681	14.29	682,494	6.50
psychiatry	60,010	979,760	16.33	792,925	13.21
psychology, multidisciplinary	54,468	761,349	13.98	1,182,361	21.71

WoS subject category	P	TCS	MCS	TRS	MRS
public administration	6,030	35,086	5.82	69,711	11.56
psychology, psychoanalysis	1,881	8,121	4.32	9,096	4.84
psychology, mathematical	1,679	14,961	8.91	28,960	17.25
psychology, experimental	22,310	310,785	13.93	637,822	28.59
radiology, nuclear medicine & medical imaging	90,673	1,224,287	13.50	964,871	10.64
rehabilitation	24,004	216,626	9.02	303,757	12.65
respiratory system	34,260	578,153	16.88	247,976	7.24
reproductive biology	19,120	295,044	15.43	140,726	7.36
rheumatology	33,279	633,881	19.05	285,228	8.57
social issues	3,940	25,806	6.55	43,599	11.07
psychology, social	19,818	263,463	13.29	508,901	25.68
social sciences, interdisciplinary	14,243	93,259	6.55	194,484	13.65
social sciences, biomedical	7,879	100,820	12.80	106,553	13.52
social work	8,277	46,178	5.58	70,012	8.46
sociology	17,086	109,286	6.40	244,152	14.29
soil science	21,587	271,570	12.58	290,853	13.47
spectroscopy	38,062	337,108	8.86	181,735	4.77
sport sciences	29,139	398,326	13.67	498,251	17.10
statistics & probability	38,405	234,780	6.11	193,902	5.05
surgery	153,238	1,714,439	11.19	841,662	5.49
telecommunications	37,624	204,333	5.43	173,522	4.61
theater	2,033	1,017	0.50	2,595	1.28
religion	8,983	7,948	0.88	22,425	2.50
toxicology	42,864	624,473	14.57	342,997	8.00
transplantation	17,393	231,563	13.31	78,518	4.51
transportation	4,234	38,225	9.03	66,662	15.74
tropical medicine	7,779	78,287	10.06	91,709	11.79
urban studies	5,554	40,873	7.36	90,811	16.35
urology & nephrology	72,736	1,072,224	14.74	357,565	4.92
veterinary sciences	46,153	331,729	7.19	491,806	10.66
peripheral vascular disease	47,271	1,055,573	22.33	387,233	8.19
virology	42,307	806,700	19.07	440,688	10.42
women's studies	5,499	32,389	5.89	49,330	8.97
zoology	36,531	341,246	9.34	675,852	18.50
mining & mineral processing	5,878	39,431	6.71	31,355	5.33
water resources	35,557	356,498	10.03	353,350	9.94
ethics	5,052	24,847	4.92	67,707	13.40
hospitality, leisure, sport & tourism	5,529	34,253	6.20	90,831	16.43
health care sciences & services	26,288	288,503	10.97	342,487	13.03
transportation science & technology	9,494	49,586	5.22	66,502	7.00
literary theory & criticism	468	274	0.59	1,103	2.36
agricultural engineering	6,775	105,398	15.56	109,308	16.13
critical care medicine	19,945	451,014	22.61	326,546	16.37
mathematical & computational biology	18,217	255,805	14.04	417,208	22.90
engineering, geological	8,026	55,637	6.93	51,360	6.40
integrative & complementary medicine	10,415	71,528	6.87	80,815	7.76
neuroimaging	6,395	154,838	24.21	211,251	33.03
gerontology	8,059	85,329	10.59	80,151	9.95
robotics	7,076	41,990	5.93	79,642	11.26
nanoscience & nanotechnology	51,160	885,582	17.31	615,556	12.03
cultural studies	4,078	11,367	2.79	39,767	9.75

WoS subject category	P	TCS	MCS	TRS	MRS
medical ethics	1,514	8,175	5.40	11,582	7.65
cell & tissue engineering	4,145	93,712	22.61	78,708	18.99
primary health care	2,601	24,447	9.40	23,322	8.97
audiology & speech-language pathology	7,259	67,980	9.37	93,324	12.86
logic	2,228	4,854	2.18	5,483	2.46

Figure 2. WoS categories with higher MRS vs. MCS (first figure) and with higher MCS vs. MRS (Second figure).



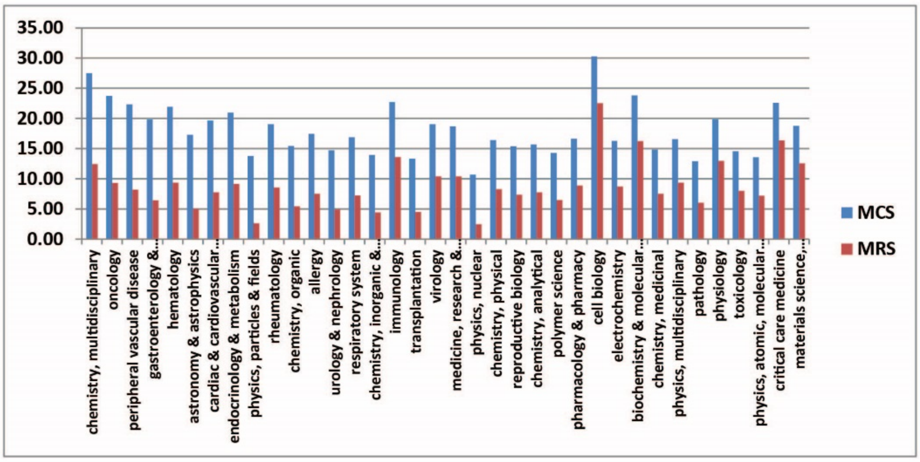
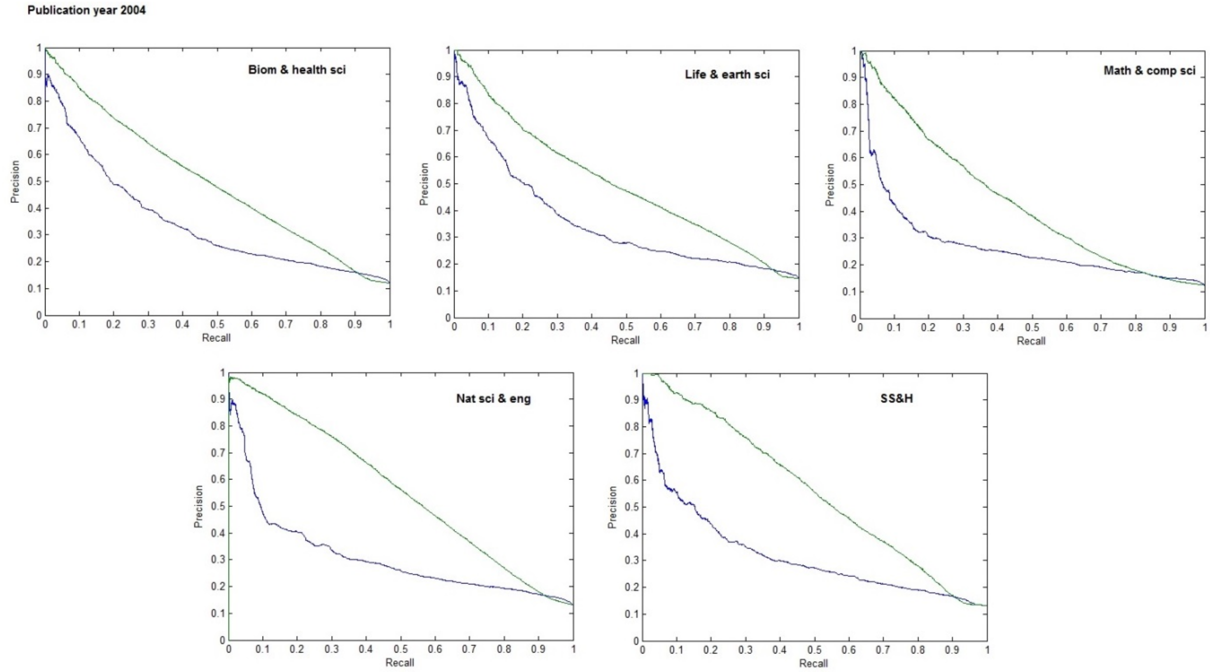


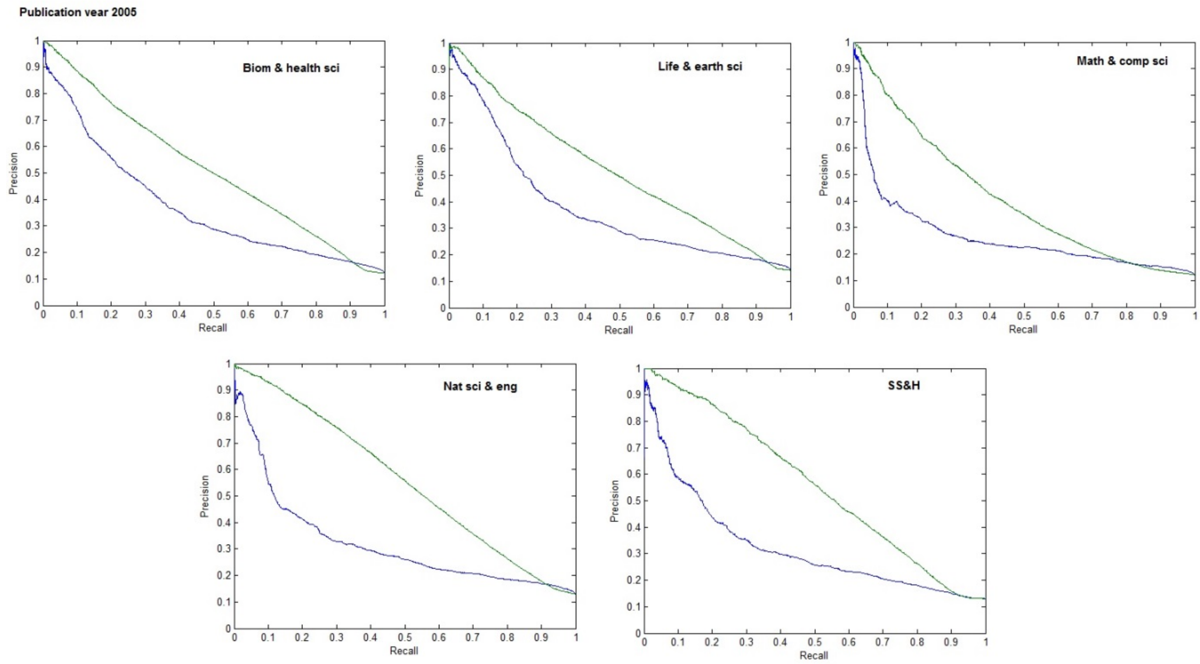
Figure 3. Precision-recall curves for JCS (blue line) and RS (green line) for identifying PPtop10% most highly cited WoS publications from the years 2003-2014 (*x* axis represents the 'Recall' and *y* axis represents the 'Precision' values).

Biom & Health Sci stands for Biomedical & health sciences; Life & Earth Sci stands for Life & earth sciecnes; Math & Comp Sci stands for Mathematical & computer sciences; Nat Sci & En stands for natural sciences & engineering; SS&H stands for Social sciences & humanities

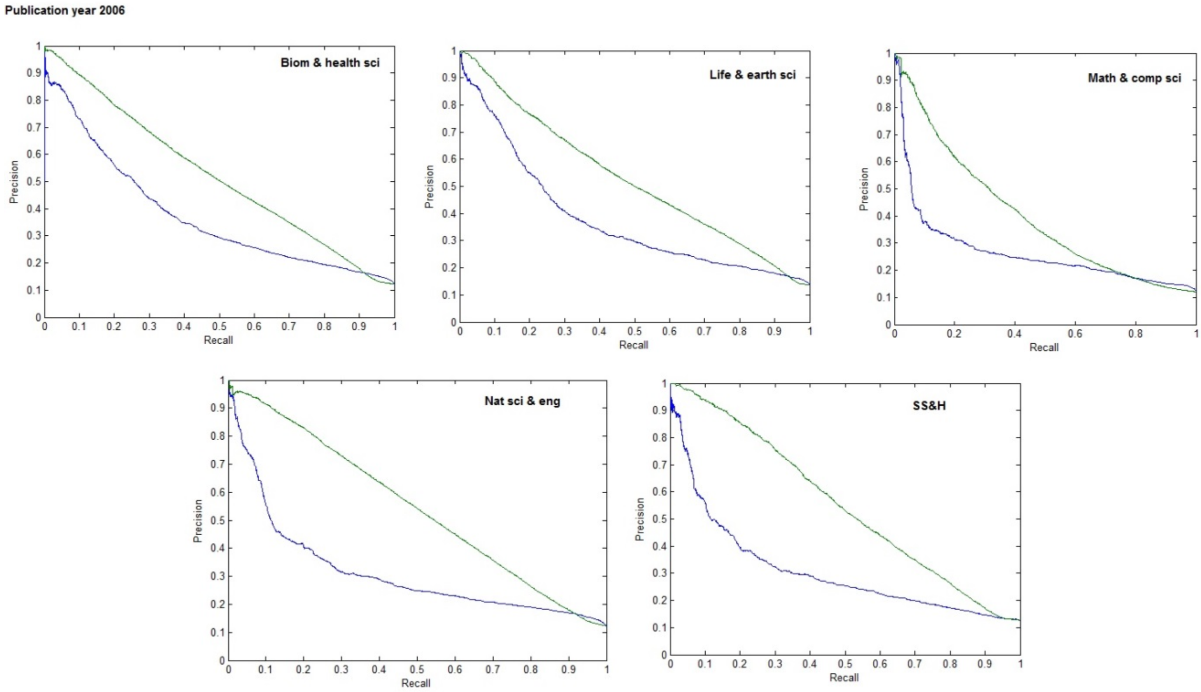
Publication year 2004



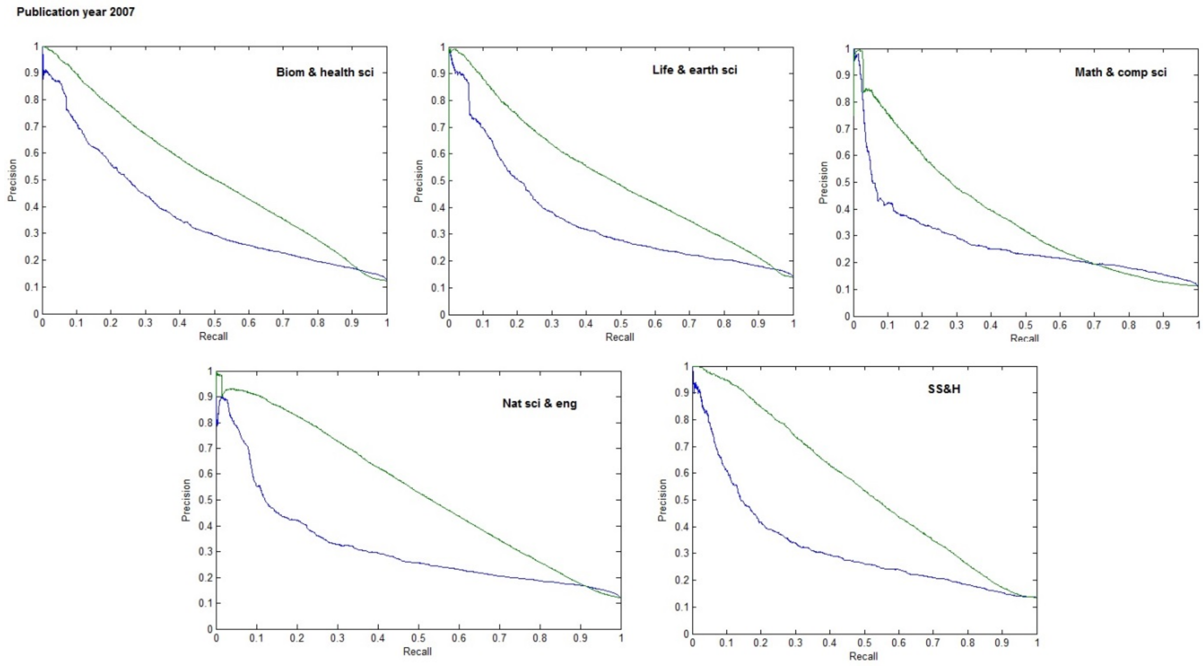
Publication year 2005



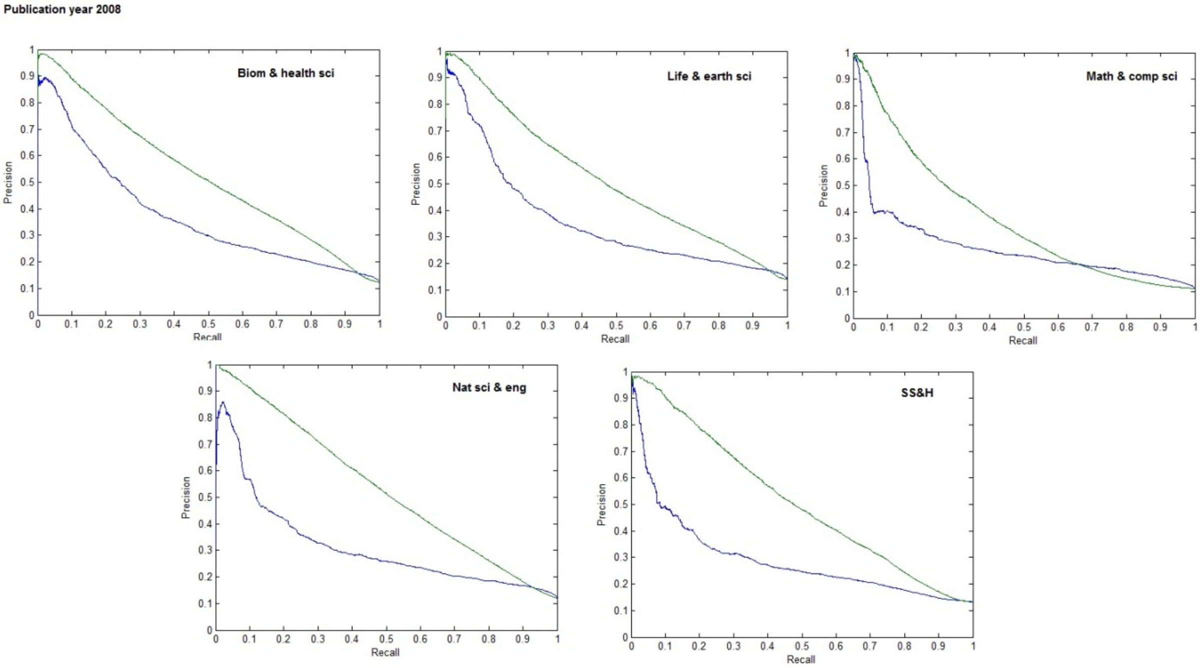
Publication year 2006



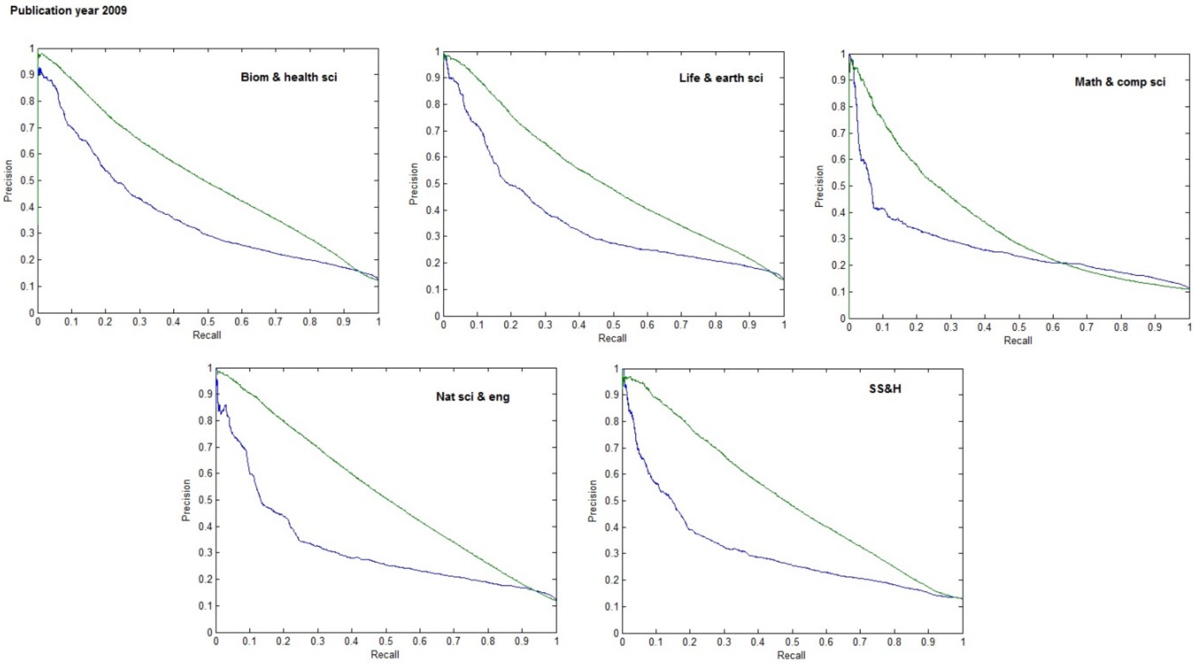
Publication year 2007



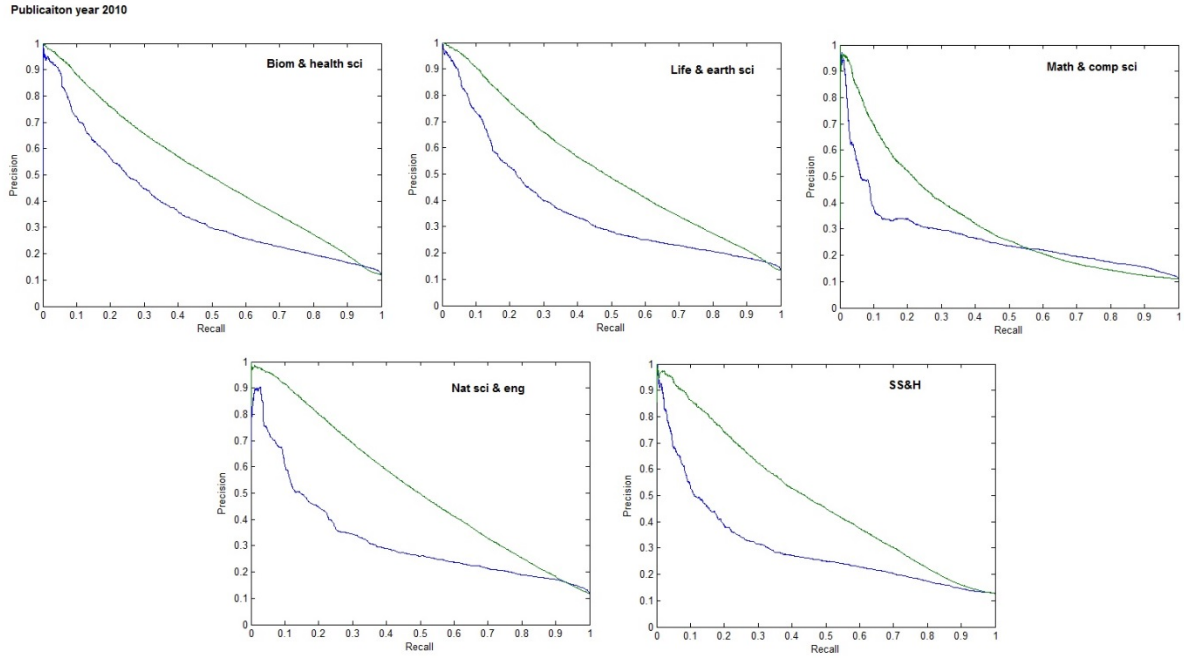
Publication year 2008



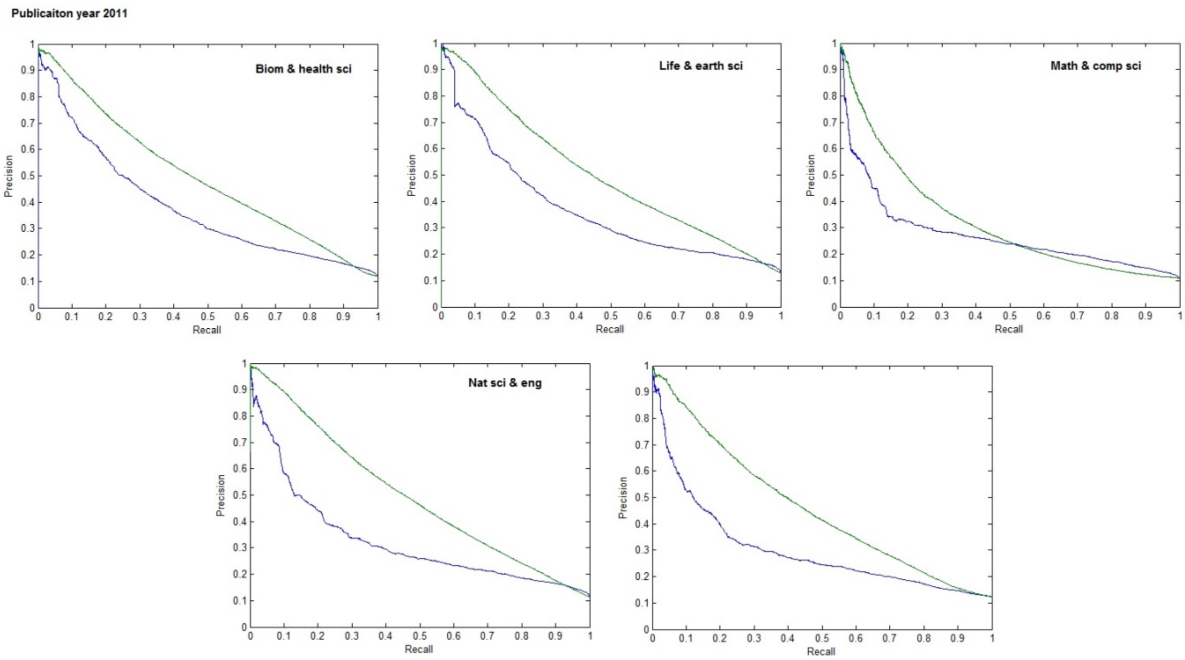
Publication year 2009



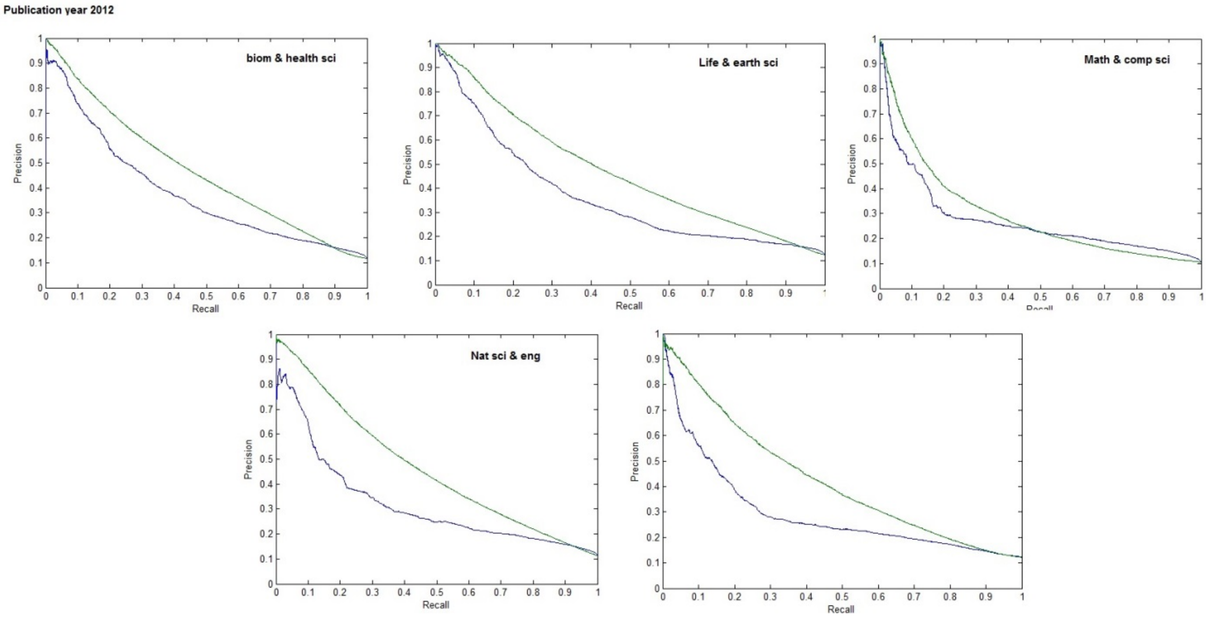
Publication year 2010



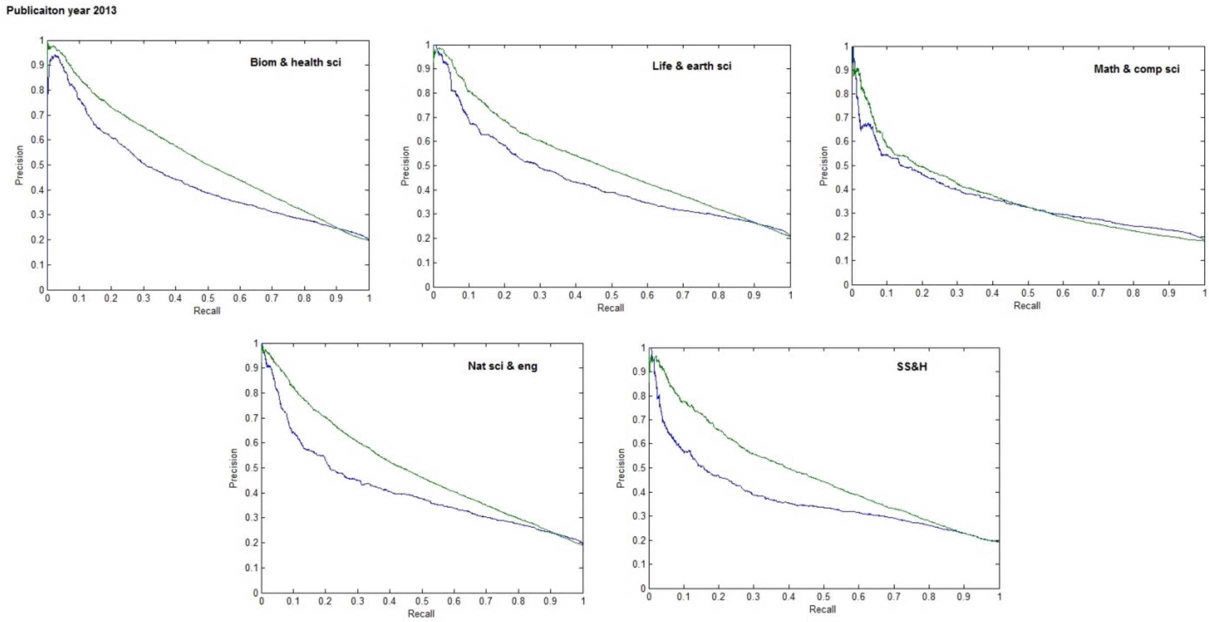
Publication year 2011



Publication year 2012



Publication year 2013



Appendix 4 (Supplementary information to Chapter 6):

Table A1. Micro-level fields that show a relatively high activity of Mendeley users classified as professors.

Micro-level field	No. readers	Terms, journals, and most read publication
295	4,287	*Terms*: portfolio; professionalism; diagnostic error; objective structured clinical examination; osce *Journals*: academic medicine; medical teacher; medical education *Publication*: ramani, s; et al. (2012). twelve tips for giving feedback effectively in the clinical environment. med teach, 34(10), 787-791.
675	3,917	*Terms*: foreign language; corrective feedback; instruction; fluency; recast *Journals*: foreign language annals; modern language journal; system *Publication*: dixon, lq; et al. (2012). what we know about second language acquisition: a synthesis from four perspectives. rev educ res, 82(1), 5-60.
466	2,844	*Terms*: politeness; talk; conversation analysis; language ideology; identity *Journals*: journal of pragmatics; discourse & society; discourse studies *Publication*: heritage, j (2012). the epistemic engine: sequence organization and territories of knowledge. res lang soc interac, 45(1), 30-52.
1344	2,172	*Terms*: student evaluation; peer assessment; feedback; doctoral education; doctoral student *Journals*: studies in higher education; teaching in higher education; assessment & evaluation in higher education *Publication*: amundsen, c; et al. (2012). are we asking the right questions?: a conceptual review of the educational development literature in higher education. rev educ res, 82(1), 90-126.
792	2,021	*Terms*: efficiency; data envelopment analysis; dea; application; china *Journals*: European journal of operational research; journal of productivity analysis; journal of the operational research society *Publication*: guan, jc; et al. (2012). modeling the relative efficiency of national innovation systems. res policy, 41(1), 102-115.

Table A2. Micro-level fields that show a relatively high activity of Mendeley users classified as researchers.

Micro-level field	No. readers	Terms, journals, and most read publication
24	9,321	*Terms*: atlantic meridional overturning circulation; eddy; indian ocean dipole; antarctic circumpolar current; seasonal climate summary southern hemisphere *Journals*: journal of climate; geophysical research letters; climate dynamics *Publication*: booth, bbb; et al. (2012). aerosols implicated as a prime driver of twentieth-century north atlantic climate variability. nature, 484(7393), 228-u110.
99	3,462	*Terms*: molecular cloud; protostar; massive star formation; young stellar object; dark cloud *Journals*: astrophysical journal; astronomy & astrophysics; monthly notices of the royal astronomical society *Publication*: glover, sco; et al. (2012). is molecular gas necessary for star formation?. mon not r astron soc, 421(1), 9-19.
1503	2,931	*Terms*: freeze; biosimilar; polysorbate; formulation; deamidation *Journals*: journal of pharmaceutical sciences; pharmaceutical research; analytical biochemistry

		Publication: liu, hc; et al. (2012). disulfide bond structures of igg molecules structural variations, chemical modifications and possible impacts to stability and biological function. mabs-austin, 4(1), 17-23.
2051	2,832	*Terms*: cho cell; chinese hamster ovary cell; mammalian cell culture; serum free medium; cho *Journals*: biotechnology and bioengineering; biotechnology progress; journal of biotechnology *Publication*: kim, jy; et al. (2012). cho cells in biotechnology for production of recombinant proteins: current state and further potential. appl microbiol biot, 93(3), 917-930.
35	2,356	*Terms*: sunspot; active region; flare; coronal mass ejection; solar flare *Journals*: astrophysical journal; solar physics; astronomy & astrophysics *Publication*: miyake, f; et al. (2012). a signature of cosmic-ray increase in ad 774-775 from tree rings in japan. nature, 486(7402), 240-242.

Table A3. Micro-level fields that show a relatively high activity of Mendeley users classified as students.

Micro-level field	No. readers	Terms, journals, and most read publication
17	66,792	*Terms*: international joint venture; export performance; radical innovation; technological capability; strategic alliance *Journals*: strategic management journal; research policy; journal of product innovation management *Publication*: garcia-morales, vj; et al. (2012). transformational leadership influence on organizational performance through organizational learning and innovation. j bus res, 65(7), 1040-1050.
15	61,774	*Terms: advertising; brand; consumer; adoption; trust* *Journals: journal of business research; psychology & marketing; journal of consumer research* *Publication: kim, aj; et al. (2012). do social media marketing activities enhance customer equity? an empirical study of luxury fashion brand. j bus res, 65(10), 1480-1486.*
22	55,808	*Terms: work family conflict; burnout; organizational citizenship behavior; leader member exchange; transformational leadership* *Journals: journal of applied psychology; leadership quarterly; journal of organizational behavior* *Publication: grant, am (2012). leading with meaning: beneficiary contact, prosocial impact, and the performance effects of transformational leadership. acad manage j, 55(2), 458-476.*
226	43,182	*Terms: corporate social responsibility; csr; environmental performance; environmental management system; business ethic* *Journals: journal of business ethics; journal of cleaner production; business ethics quarterly* *Publication: herazo, b; et al. (2012). sustainable development in the building sector: a canadian case study on the alignment of strategic and tactical management. proj manag j, 43(2), 84-100.*
222	33,912	*Terms: sigma; supply chain integration; tqm; mass customization; manufacturing strategy* *Journals: industrial marketing management; international journal of production economics; international journal of production research* *Publication: golicic, sl; et al. (2012). implementing mixed methods research in supply chain management. int j phys distr log, 42(8-9), 726-741.*

Table A4. Micro-level fields that show a relatively high activity of Mendeley users classified as librarians.

Micro-level field	No. readers	Terms, journals, and most read publication
175	2,337	*Terms*: bibliometric analysis; h index; impact factor; citation analysis; citation *Journals*: scientometrics; journal of the american society for information science and technology; journal of informetrics *Publication*: li, xm; et al. (2012). validating online reference managers for scholarly impact measurement. scientometrics, 91(2), 461-471.
277	1,787	*Terms*: information literacy; user; clef; academic library; web search *Journals*: journal of the american society for information science and technology; lecture notes in computer science; journal of academic librarianship *Publication*: schilling, k; et al. (2012). best methods for evaluating educational impact: a comparison of the efficacy of commonly used measures of library instruction. j med libr assoc, 100(4), 258-269.
768	1,182	*Terms*: knowledge translation; research utilization; nurse; clinical guideline; nursing practice *Journals*: implementation science; journal of evaluation in clinical practice; journal of the medical library association *Publication*: lasserre, k (2012). expert searching in health librarianship: a literature review to identify international issues and australian concerns. health info libr j, 29(1), 3-15.
228	820	*Terms*: nurse practitioner; job satisfaction; nurse staffing; patient outcome; retention *Journals*: nurse education today; journal of advanced nursing; journal of nursing administration *Publication*: hayes, lj; et al. (2012). nurse turnover: a literature review - an update. int j nurs stud, 49(7), 887-905.
12	741	*Terms*: year old child; childhood overweight; preschool child; physical education; adolescence *Journals*: public health nutrition; bmc public health; international journal of obesity *Publication*: brown, hw; et al. (2012). exploring the factors contributing to sibling correlations in bmi: a study using the panel study of income dynamics. obesity, 20(5), 978-984.

Table A5. Micro-level fields that show a relatively high activity of Mendeley users classified as other professionals.

Micro-level field	No. readers	Terms, journals, and most read publication
66	4,387	*Terms*: wolf; canis; white tailed deer; elk; brown bear *Journals*: journal of wildlife management; wildlife society bulletin; biological conservation *Publication*: foster, rj; et al. (2012). a critique of density estimation from camera-trap data. j wildlife manage, 76(2), 224-236.
967	3,078	*Terms*: tursiops truncatus; bottlenose dolphin; dolphin; cetacea; whale *Journals*: marine mammal science; journal of the acoustical society of america; marine ecology progress series *Publication*: rolland, rm; et al. (2012). evidence that ship noise increases stress in right whales. p roy soc b-biol sci, 279(1737), 2363-2368.
132	2,438	*Terms*: helical tomotherapy; stereotactic body radiotherapy; image; volumetric modulated arc therapy; imrt

		Journals: medical physics; physics in medicine and biology; international journal of radiation oncology biology physics *Publication*: bissonnette, jp; et al. (2012). quality assurance for image-guided radiation therapy utilizing ct-based technologies: a report of the aapm tg-179. med phys, 39(4), 1946-1963.
435	2,259	*Terms*: warfarin; new oral anticoagulant; rivaroxaban; venous thromboembolism; stroke prevention *Journals*: thrombosis and haemostasis; thrombosis research; journal of thrombosis and haemostasis *Publication*: douketis, jd; et al. (2012). perioperative management of antithrombotic therapy antithrombotic therapy and prevention of thrombosis, 9th ed: american college of chest physicians evidence- based clinical practice guidelines. chest, 141(2), e326s-e350s.
242	2,214	*Terms*: proximal humeral fracture; total shoulder arthroplasty; rotator cuff repair; adhesive capsulitis; arthroscopic repair *Journals*: journal of shoulder and elbow surgery; arthroscopy-the journal of arthroscopic and related surgery; american journal of sports medicine *Publication*: kibler, wb; et al. (2012). scapular dyskinesis and its relation to shoulder injury. j am acad orthop sur, 20(6), 364-372.

List of publications

International peer reviewed journal papers:

Zahedi, Z., Costas, R. & Wouters, P. (2014). How well developed are Altmetrics? Cross disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 101(2): 1491-1513. DOI: 10.1007/s11192-014-1264-0

Zahedi, Z., Costas, R. & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the American Society for Information Science and Technology*. 68 (10): 2511-2521. DOI: 10.1002/asi.23883

Zahedi, Z., & Haustein, S. (2018). On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications. Journal of Informetrics, 12 (1): 191-202. DOI: 10.1016/j.joi.2017.12.005

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. PloS one, 13(5), e0197326. DOI: 10.1371/journal.pone.0197326

Zahedi, Z., and van Eck, Nees. (Accepted). Exploring topics of interest of Mendeley users. Journal of Altmetrics.

Costas, R., **Zahedi, Z.** & Wouters, P. (2015). The thematic orientation of publications mentioned on social media: large-scale disciplinary comparison of social media metrics with citations, *Aslib Journal of Information Management*, 67(3): 260-288. DOI: 10.1108/AJIM-12-2014-0173

Costas, R., **Zahedi, Z.** & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the American Society for Information Science and Technology*, 66 (10): 2003-2019. DOI: 10.1002/asi.23309

Robinson-García, N., Torres-Salinas, D., **Zahedi, Z.** & Costas, R. (2014). New data, new possibilities: Exploring the insides of Altmetric.com. *El Profesional de la información*, 23(4): 359-366.

Scientific report:

Costas, R., Meijer, I., **Zahedi, Z.,** Wouters, P. (2013). *The Value of Research Data Metrics for datasets from a cultural and technical point of view*. Denmark: Knowledge-Exchange.

Book chapters:

Wouters, P., **Zahedi, Z.**, & Costas, R. (2018). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), Handbook of Quantitative Science and Technology Research. Springer.

Altmetric data quality code of conduct (2016). Outputs of the *NISO Alternative Assessment Metrics Project: a recommended practice of National Information Standards Organization*. Published by the National Information Standards Organization, ISBN: 9781937522711.

Invited talks:

Zahedi, Z., *"Challenges of altmetrics data"*, COAR metrics workshop, 14-15 May 2018, Hamburg, Germany.

Zahedi, Z., *"Altmetric divide: on the challenges of the imbalance use of scholarly social media platforms across different countries",* the First Altmetric conference, *13 January 2018,* Shahid Beheshti University, Tehran, Iran.

Zahedi, Z., *"Open Science and Altmetrics:* challenges and opportunities", The BE-OPEN (Boosting Engagement of Serbian Universities in Open Science) workshop, 17-19 May 2017, Leiden, The Netherlands.

Zahedi, Z., "A*ltmetrics: opportunities or risk*?*"* at the information day "Bibliometrics, Scientometrics & Alternative metrics: which tools for which strategies?", Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU), 1st April 2015, BULAC, France, Paris.

Zahedi, Z., "*Introduction to Altmetrics"*, 12 March 2014, Armook Educational Group & Shahid Beheshti University, Tehran, Iran.

Zahedi, Z., "*Research Evaluation: Using Altmetrics*", 24 October 2013, Maastricht University Library, Maastricht, The Netherlands.

International peer reviewed Conference papers:

Full conference papers

Zahedi, Z., Costas, R. (2018). Challenges in the quality of social media data across altmetric data aggregators. In Proceedings of the 23rd *International Conference in Science & Technology Indicators* (STI), 12-14, September 2018, Leiden, the Netherlands.

Zahedi, Z., Costas, R. (2017). How visible is the research of different countries? an analysis of global vs. local reach of WoS publications on Twitter. In Proceedings of the *16th International Conference on Scientometrics and Informetrics* (ISSI), 16-20 October 2017, Wuhan University, Wuhan, China.

Costas, R., van Honk, J., Calero-Medina, C., **Zahedi, Z.** (2017). Exploring the descriptive power of altmetrics: case study of Africa, USA and EU28 countries (2012-2014). In Proceedings of the *22ed International Conference in Science & Technology Indicators* (STI), 6-8 September 2017, Paris, France.

Zahedi, Z., Costas, R., Larivière, V., Haustein, S. (2016). What makes papers visible on social media? An analysis of various document characteristics. In Proceedings of the *21th International Conference in Science & Technology Indicators* (STI), 13-16, September, 2016, Valencia, Spain.

Zahedi, Z., & Van Eck, N. (2015). Identifying topics of interest of Mendeley users using the text mining and overlay visualization functionality of VOSviewer. In Proceedings of the *20th International Conference in Science & Technology Indicators* (STI), 2-4, September, 2015, Lugano, Switzerland.

Van der Weijden, I., Nane, T., Costas, R., **Zahedi, Z.,** & Meijer, I. (2015). The importance of personal grants on the scientific performance of early career scholars: A Dutch case study. In Proceedings of the *20th International Conference in Science & Technology Indicators* (STI), 2-4, September, 2015, Lugano, Switzerland.

Zahedi, Z., Costas, R., & Paul Wouters (2015). Do Mendeley Readership Counts Help to Filter Highly Cited WoS Publications Better than Average Citation Impact of Journals (JCS)? In Proceedings of the *15th International Conference on Scientometrics and Informetrics* (ISSI), 29 Jun-4 July, 2015, Bogazici University, Istanbul (Turkey).

Zahedi, Z., Costas, R., & Paul Wouters (2014). Broad altmetric analysis of Mendeley readerships through the career stages of the readers. In Proceedings of the *19th International Conference in Science & Technology Indicators* (STI), 3-5, September, 2014, Leiden University, The Netherlands.

Zahedi, Z., Costas, R., & Paul Wouters (2014). Assessing the impact of the publications read by the different Mendeley users: Is there any different pattern among users?. In Proceedings of the *35th International Association of Technological University Libraries* (IATUL), 2-5 June 2014, Aalto University, Helsinki, Finland.

Van der Weijden, I., **Zahedi, Z.** , Belder, R., & Meijer, I. (2014). Researchers with a personal grant: perceptions on societal relevance and outputs. Presented at the Research Funding & Dynamics of Science workshop, The Research Network Sociology of Science and Technology Network (SSTNET) of the European Sociological Association (ESA) and the Centre for Science and Technology Studies (CWTS), 12-13 June, 2014, Leiden University, The Netherlands.

Van der Weijden, I., **Zahedi, Z.**, Belder, R., Must, U, & Meijer, I (2014). Gender differences in societal orientation and output of individual scientists. In Proceedings of the *19th International Conference in Science & Technology Indicators* (STI), 3-5, September, 2014, Leiden University, The Netherlands.

Zahedi, Z. (2014). The use of English language Iranian international publications by Mendeley users. In Proceedings of the *first national conference on Scientometrics*, 21-22 May 2014, Isfahan University, Isfahan, Iran.

Short conference and workshop papers and presentations

Zahedi, Z. (2018). Characterizing scholarly Twitter users and their interactions across different countries. The 2018 Altmetrics workshop, 25 September 2018, London, UK.

Zahedi, Z. (2017). What explains the imbalance use of social media across different countries? A cross country analysis of presence of Twitter users tweeting scholarly publications. *The 4:AM conference*, 26-29 September 2017, Toronto, Canada.

Zahedi, Z., Costas, R., Larivière, V., Haustein, S. (2016). On the relationships between bibliometric and altmetric indicators: the effect of discipline and density level. *The 2016 Altmetrics Workshop*, 27 September 2016, Bucharest, Romania.

Costas, R., Haustein, S., **Zahedi, Z.**, Larivière, V. (2016). Exploring paths for the normalization of altmetrics: applying the Characteristic Scores and Scales. *The 2016 Altmetrics Workshop*, 27 September 2016, Bucharest, Romania.

Fraumann, G., Mugnaini, R., Costas, R., Packer, A.L., **Zahedi, Z**. (2016). Is there an 'altmetric advantage' for SciELO Brazil Journals that are active in Twitter?. *The 2016 Altmetrics Workshop*, 27 September 2016, Bucharest, Romania.

Costas, R., Van Honk, J., **Zahedi, Z.**, Calero-Medina, C. (2016). Discussing practical applications for altmetrics: development of 'social media profiles' for African, European and North American publications. *The 3:AM conference*, 28-29 September 2016, Bucharest, Romania.

Zahedi, Z., Fenner, M., & Costas, R. (2015). Consistency among altmetrics data provider/aggregators: what are the challenges? *The 2015 Altmetrics workshop*, 9 October 2015, Amsterdam Science Park, Amsterdam, the Netherlands.

Zahedi, Z., Fenner, M., & Costas, R. (2015). Challenges in altmetric data collection: are there differences among different altmetric providers/aggregators? *The 2:AM conference*, Amsterdam Science Park, 7-8, October 2015, Amsterdam, the Netherlands.

Fraumann, G., **Zahedi, Z.,** & Costas, R. (2015). What do we know about Altmetric.com sources? A study of the top 200 blogs and news sites mentioning scholarly output. *The 2015 Altmetrics workshop*, 9 October 2015, Amsterdam Science Park, Amsterdam, the Netherlands.

Zahedi, Z., Haustein, S., & Bowman, Tim D (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. *The Metrics14: ASIS&T Workshop on Informetrics and Scientometrics Research*, October 31-November 5 2014, Seattle, WA.

Zahedi, Z., Fenner, M., & Costas, R (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs. *The 2014 Altmetrics workshop and ACM Web Science Conference 2014*, 23-26 June 2014, Indiana University, Bloomington, US.

Zahedi, Z., & Van Eck, N (2014). Visualizing readership activity of Mendeley users using VOSviewer. *The 2014 Altmetrics workshop and ACM Web Science Conference 2014*, 23-26 June 2014, Indiana University, Bloomington, US.

Costas, R., **Zahedi, Z.** & Wouters, P. (2014). Disentangling the meaning of 'altmetrics': content analysis of Web of Science scientific publications. *The 2014 Altmetrics workshop and ACM Web Science Conference 2014*, 23-26 June 2014, Indiana University, Bloomington, US.

Zahedi, Z., Costas, R., & Paul Wouters (2013).What is the impact of the publications read by the different Mendeley users? Could they be considered as signals of future impact? *PLOS ALM Workshop*, 9-12 Oct, San Francisco, CA.

Acknowledgments

I have a great deal of affection for all who have helped me to shape my academic life. As I look back, I see there were so many people who contributed directly or indirectly to the completion of this thesis that without their help and support it would have been almost impossible to conduct this PhD project. I hereby want to thank you all.

I would like to express my sincere and deep gratitude to my supervisor Prof. Paul Wouters and my co-supervisor Dr. Rodrigo Costas for all their thoughtful guidance throughout my PhD research. Paul, I am very grateful for the opportunity you gave me to carry out my PhD research at CWTS. Thanks for giving me the freedom to pursue this line of research and for all your critical opinions and valuable advice, which helped me to improve my critical thinking. Rodrigo, words fall short to express my deep thanks to you. You were always there to support, encourage, and guide me. I really appreciate your patience and understanding.

I wish to thank all my colleagues at CWTS. My special thanks goes to Prof. Cornelis van Bochove for all his kind advice and Prof. Ludo Waltman, Dr. Ingeborg Meijer, and Dr. Inge van der Weijden for all their help and support. Particular thanks go to Henri de Winter and Erik van Wijk. Without their great support for the altmetrics data collection I could never have done my research. Many thanks to Wolfgang, Julia, Andrea, Alfredo, Lidia, Suze, Anne Marie, and Petra. Thanks to Wout for his help with the Dutch translation of the propositions. Special thanks to Andrea for her great help in designing the cover of my thesis. Heartfelt thanks to Clara for all her emotional care and support. Cathelijn, I am so lucky to have you as my best friend and I am delighted that you have agreed to act as my paranimf. Thanks for all your kindness and particularly for your great help with the Dutch translation of the summary of my thesis. Marcia, it is my great pleasure to have you too as my paranimf. Joost, I wish to thank you too for all your thoughtful advice about everything related to living and working in the Netherlands. Particular thanks to Prof. Mike Thelwall (University of Wolverhampton), Dr. Diana Maynard and Dr. Genevieve Gorrell (University of Sheffield), Prof. Vincent Larivière, and Dr. Stefanie Haustein (University of Montreal) for hosting me at their universities. Many thanks to the altmetric data providers for all their practical support. I would like to express my gratitude to the Ministry of Science and Technology of Iran for granting me the PhD scholarship. I am very grateful to Prof. Jahangiri, Iranian ambassador to the Netherlands, from whom I was honored to receive an honorable mention for my scientific activities. I would like to express my sincere gratitude to my dear colleagues at Persian Gulf University, particularly Mrs. Alavi, Dr. Pakizeh, Dr. Tajik, Dr. Karimi, Dr. Sarrafzadeh, Dr. Nemati, Dr. Movahhed, Dr. Ghahramani, Mr. Nakisa, Mr. Akmiul, Mr. Khorshidian, Dr. Ghasemiyeh, and Dr. Mosleh for all their substantial support. I'm also grateful to all my Iranian friends in the Netherlands whose support have made me feel at home. Special thanks go to all members of Zahedi & Izadkhah family, particularly my parents and mother- and father-in-law for all their support during my stay in the Netherlands. Last but not least, I am forever indebted to my dear husband for all his unlimited love, support, and kindness throughout all these years. Mohammad, thank you immeasurably for all you did to allow me to pursue my scientific goals.

Curriculum Vitae

Zohreh Zahedi was born on February 22, 1980 in city of Lar (Iran). After her birth, she moved with her parents to Shiraz. Zohreh received her Diploma in Experimental Sciences (Biology) with distinction, from Mehraeen High school in Shiraz in the year 1998. She obtained her Bachelor Degree in Knowledge & Information Science with distinction, from Shiraz University in the year 2003. She ranked 2nd among the 2053 candidates in the national university entrance exam for a master program in Knowledge &Information Science in 2003 and hence was recognized as an 'exceptional talent' by the National Organization of Educational Testing of Iran. In the year 2006, she received her Master Degree in Knowledge & Information Science with distinction, from Shiraz University. Zohreh has 6 years of both academic and professional experience after her graduation. From 2006 to 2008, she was the head of the department of international collection development at the Regional Information Center for Science & Technology in Shiraz. From 2008 to 2012, she worked as a faculty member and university lecturer at the department of Information Science of the Persian Gulf University in Bushehr. In this position she received the 'excellent university lecturer' award from the dean of the university for her contribution to the development of the Information Science undergraduate program. In 2012, she was awarded a competitive PhD scholarship by the Ministry of Science, Research & Technology of Iran to do a PhD program abroad. Zohreh did her PhD studies at the Centre for Science and Technology Studies (CWTS) of Leiden University from January 2013 until February 2018. She was awarded the '*Honorable mention' of the prestigious 'Eugene Garfield Doctoral Dissertation Scholarship Award*' by the International Society for Scientometrics and Informetrics committee for her PhD dissertation. As a result of her thesis which focuses on the possibilities and challenges of social media metrics for informing research evaluation, Zohreh has contributed to the development of altmetric data infrastructure and databases at CWTS. Beyond the immediate subject of her PhD, she was also involved in some altmetrics and bibliometrics projects at CWTS. She has also been involved in teaching 'Social media metrics' as part of the 'Measuring science and research performance' course offered by CWTS. Zohreh is currently a member of the 'Applied Scientometric lab' and the 'Quantitative Science Studies' research group and contributes particularly to the 'Social media metrics' research line at CWTS. Besides, she contributed to the development of the US National Information Standard Organization code of conduct for altmetric data in 2016. She served as the project advisor of the TU Delft Library altmetric project from October 2017 to September 2018. She has published a number of papers on the topic of social media metrics in the leading international peer reviewed journals and in conference proceedings. Zohreh has also served as reviewer and scientific committee member of prestigious journals and conferences in the field and has supervised some undergraduate and graduate students' theses.