Understanding the value of social media metrics for research evaluation

Zahedi, Z.

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*.