Scientometrics
https://doi.org/10.1007/s11192-018-2674-1

Cross-metric compatability and inconsistencies of altmetrics

Christine Meschede[1] · **Tobias Siebenlist**[1]

Received: 26 July 2017

Abstract Metrics like the number of tweets or Mendeley readers are currently discussed as an alternative to evaluate research. These alternative metrics (altmetrics) still need to be evaluated in order to fully understand their meaning, their benefits and limitations. While several preceding studies concentrate on correlations of altmetrics with classical measures like citations, this study aims at investigating metric-compatibility within altmetrics. For this purpose, 5000 journal articles from six disciplines have been analyzed regarding their metrics with the help of the aggregators PlumX and Altmetric.com. For this set, the highest numbers of events have been recognized regarding Mendeley readers, followed by Twitter and Facebook mentions. Thereby variations considering the aggregators could be observed. Intra-correlations between the metrics across one aggregator have been calculated, as well as inter-correlations for the corresponding metrics across the aggregators. For both aggregators, low to medium intra-correlations could be calculated which shows the diversity of the different metrics. Regarding inter-correlations, PlumX and Altmetric.com are highly consistent concerning Mendeley readers ($r = 0.97$) and Wikipedia mentions ($r = 0.82$), whereas the consistency concerning Twitter ($r = 0.49$), blogs ($r = 0.46$) and Reddit ($r = 0.41$) is on a moderate level. The sources Facebook ($r = 0.29$), Google+ ($r = 0.28$) and News ($r = 0.11$) show only low correlations.

Keywords Altmetrics · Aggregators · Correlations · Cross-metric compatibility · Inconsistencies · Mendeley

✉ Christine Meschede
christine.meschede@hhu.de

[1] Department of Information Science, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Published online: 17 February 2018

Introduction

With the emergence of more and more social media portals and tools which scientists can use to present their research and collaborate with others, the scientific world is moving online. To name just a few examples, publications are shared on ResearchGate, mentioned in tweets on the microblogging system Twitter or collected via the reference manager Mendeley. According to Nicholas and Rowlands (2011), researchers use social media tools in different stages of their work, for discovering interesting research opportunities but also to share their findings with others. Open science is an important concept in this context, trying to make research and data accessible to a broader audience. Along with the movement towards online scholarly communication and open science, comes the possibility to quantify anything we find on the web. Referring to the aforementioned examples, the number of tweets for a publication or the amount of users saving an article on Mendeley can be counted, today primary known as altmetrics.

At the same time, researchers are strongly under pressure to prove their contribution to society and quantitative indicators—also in the form of altmetrics—are one way to do so. Hence it is essential that these measures fulfill criteria like transparency and robustness (Martin 2015). As the meaning and benefits of altmetrics are not yet fully understood, the work presented here aims at expanding current research on altmetrics. It will be investigated how consistent certain metrics are by calculating correlations between them. In the following, an introduction into the background of the topic will be given. The current state of the art on altmetric research is depicted, including altmetric aggregators and the necessity to evaluate altmetrics.

Definition of altmetrics

Several metrics have been produced in the past, mostly based on citations, which have long been the accepted indicator for measuring the impact of an author or a journal. Whether it is the h-Index as an author-level metric that quantifies the individual's research output (Hirsch 2005) or the Journal Impact Factor (JIF) for assessing journals (Garfield 1955), these indicators rely on citations. The amount of citations an author receives thereby does not seem to be arbitrary. Merton (1988) emphasizes that famous researchers get credit for their work aplenty while relatively new ones do only receive little reputation for similar works. This phenomenon is known as the *Matthew Effect* and can also be applied to popular journals. Beside this, there is one important disadvantage of using citations for measuring impact—it is an indicator that needs its time. From publication to citation several years can pass as the work has to be read, embedded in a new publication and then this new work has to be published as well (Sud and Thelwall 2014).

With the scientific world moving online, giving credit to articles and authors does not only take place in a formal publication but also in social networks or blogs. Hence, not only citations can be counted, but also all these online events. In contrast to citations this is a fast process, as a tweet can be written directly after or even while reading an article. A bunch of new metrics has risen out of this development. As they differ from classical citation counting in scientometrics, the new term altmetrics has been proposed by Priem (2010) via a hashtag in a tweet. Although the term has been introduced as recently as 2010, a similar concept already existed before in the form of *article-level metrics*. This approach includes all measures for an individual article including social media events, but does not cover further research objects and hence "fails to imply *diversity* of measures" (Priem 2010). Even older is the idea of *webometrics*, quantitative measures of network-based communication (Almind and Ingwersen 1997). Although parallels to altmetrics can be drawn, Thelwall (2016, pp. 342–343) delineates these concepts since “altmetrics had the potential to be used by *publishers* to help guide readers to the most important recently published articles, which never seemed be a realistic goal for webometrics” and because “altmetrics has not produced any clearly named theories”.

This alone does not define the term altmetrics. Later in 2010, Priem et al. (2010, para. 1) extended their idea of altmetrics in the altmetrics manifesto: “the growth of new, online scholarly tools allows us to make new filters; these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem”. In other definitions the focus is set on social media, e.g., Weller (2015, p. 262) specifies altmetrics as “evaluation methods based on various user activities in social media environments”. Hence, there is not yet a common understanding of altmetrics, whereas there is only a consensus in the fact that altmetrics describe alternatives to ‘traditional’ metrics like citations (Erdt et al. 2016; Haustein et al. 2015). Against this background, Haustein et al. (2015, p. 373) further propose a generic definition of 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”. In this work, this broader understanding of altmetrics will be considered for all further analyses.

Altmetric aggregators

Altmetric providers aim at presenting its users a variety of metrics for publications ranging from classical citation data to usage counts and social media mentions. Some of the services directly collect data from sources like Twitter or Mendeley (primary aggregators), others serve as secondary aggregators by relying on other existing altmetric providers (Erdt et al. 2016). All of these aggregators provide the user with a different set of metrics, often including also citation and usage metrics. Thereby each aggregator has its own strategy in presenting the metrics to the user. Altmetric.com e.g., provides metrics for a wide range of publications. A user can freely access the metrics for any publications by a bookmarklet or open API. The for-profit provider gathers data from a variety of sources from policy documents, blogs and citations to reference managers and social media. The special feature of Altmetric.com is that an aggregated score is calculated from the metrics, called the Altmetric Attention Score which is visualized in form of an “altmetric donut”. This score is a weighted count based on the weightings in Table 1. It is argued that a news article brings more attention to a research output than a tweet and hence news are weighted higher. Additionally, some data sources like Mendeley or CiteULike are not taken into account for calculating the score which is explained as follows: “we can’t display the actual profiles, and we want all our data to be fully auditable” (Altmetric 2016a, para. 8). Mukherjee et al. (2018) warn against the use of the Altmetric Attention Score and similar aggregated measures because of the arbitrariness such indicators entail.

Another for-profit aggregator is Plum Analytics which was founded in 2012. In 2014 Plum Analytics became part of EBSCO Information Services and has been acquired by Elsevier in 2017 (Plum Analytics 2017). Different products are offered, including PlumX Metrics which provides metrics from the categories *usage*, *captures*, *mentions*, *social media*, and *citations*. The products are usually only offered for paying institutions and are customized for them, but upon request researchers may obtain access to a subset of the data. The aggregators provide different metrics from several sources, PlumX e.g., is the only one collecting data from EBSCO due to the fact that Plum Analytics was part of it. Moreover, CiteULike mentions and LinkedIn are no longer tracked by Altmetric.com

Table 1 Weightings of the Altmetric Attention Score (Altmetric 2016a)

Data source	Weighting
News	8
Blogs	5
Twitter	1
Facebook	0.25
Sina Weibo	1
Wikipedia	3
Policy Documents (per source)	3
Q&A	0.25
F1000/Publons/Pubpeer	1
YouTube	0.25
Reddit/Pinterest	0.25
LinkedIn	0.5
Open Syllabus	1
Google+	1

(Altmetric 2017). PlumX just started to include CiteULike mentions in October 2017 (Lohr 2017). Other, more well-known metrics like tweets, Mendeley readers or Facebook events are covered by all of the data aggregators.

Current state of altmetric research

One of the first studies in altmetric research was the application of social bookmarking data from CiteULike, Connotea and BibSonomy for the evaluation of journals, conducted by Haustein and Siebenlist (2011). Back then, in 2011, altmetrics was a rather new topic with only a handful of publications on the topic. This has changed immensely during the last years as the number of publications is growing rapidly. Erdt et al. (2016) give an overview on the altmetric landscape and present a systematic review on existing altmetric literature performed by a meta-analysis. The topics that are covered in the 177 reviewed publications range from calculations of correlations between the different metrics (with a focus on correlations between altmetrics and citations) to an examination of user motivations in social media use.

A lot of studies concentrate on cross-metric validations and thereof most analyses focus on the calculation of correlations between citations and altmetrics (Erdt et al. 2016). In 2015, Bornmann (2015) performed a meta-analysis which aimed at aggregating results from different studies on correlations of altmetrics with classical citation counts. Thereby he compared data on the microblogging platform Twitter, the online reference managers Mendeley and CiteULike and blogs. Overall, there is no correlation of the number of tweets and traditional citation counts ($r = 0.003$). Considering the online reference managers, the correlation to citation counts provided higher values. CiteULike produces an overall correlation value of $r = 0.23$ and Mendeley a value of even $r = 0.51$. Finally, the comparison of blog posts and citations result in a low correlation of $r = 0.12$. Similarly, Erdt et al. (2016) detected that most studies reported a low correlation of altmetrics with citations whereby the highest values have been achieved by Mendeley (on average $r = 0.37$ respectively $r = 0.55$ when considering only non-zero counts).

Another popular topic is the coverage of altmetric data sources like Mendeley, Twitter or CiteULike (Erdt et al. 2016). Haustein et al. (2014) conducted an analysis on the coverage on Mendeley and CiteULike of over a thousand publications from bibliometricians. The authors identify a high coverage on Mendeley of 82% and only a low coverage on CiteULike of 28%. Zahedi et al. (2014) examined the metrics of nearly 20,000 Web of Science publications covered by the aggregator ImpactStory. Again, Mendeley is identified as the most prevalent data source with 62.6% of the publications having at least one Mendeley reader. PubMed citations were discovered for 37.4% of the publications, Twitter, Wikipedia and Delicious however only have a low coverage of 1.6, 1.4 and 0.3%. Erdt et al. (2016) reviewed 25 publications considering the coverage of eleven altmetric data sources. Overall they report a moderate coverage of Mendeley (59.2%) and Twitter (24.3%). CiteULike, Facebook, F1000 and Research Blogging achieve values of 10.6, 7.7, 6.1 and 5.5%. The remaining data sources Google+, Reddit, news, Wikipedia and blogs showed only a low overall coverage of 1.1–3.7%.

There are further studies examining the general coverage of publications in altmetric aggregators. Torres-Salinas et al. (2016) conducted an analysis on the publications of four Spanish universities and their coverage on the aggregator Altmetric.com with regard to disciplinary and institutional variations. They discovered a low overall coverage of 36%, with the highest percentage for the broad field Science, with variations considering the four institutions (24, 34, 36 and 70%). Other investigations focus on disciplinary variations of concrete altmetric data sources. Holmberg and Thelwall (2014) examine scholarly communication on Twitter with regard to a more specific classification into ten disciplines. It is noticed that in Biochemistry the highest amount of re-tweets is used (42%) and Economics shares the most links (38%).

Further, the authors report disciplinary differences in the amount of tweets classified as scholarly communication. In Biochemistry, 34% of the tweets were dedicated to this class, while in Astrophysics the value sums up to 23%. In Economics only 7% of the tweets can be considered as scholarly communication and the lowest value has been recognized for Sociology (0.5%). However, the highest amount of discipline-relevant tweets seems to come from Economics (51.5%), while in other disciplines the value varies between 4.5 and 22% (Holmberg and Thelwall 2014). Beside the variety of subtopics on altmetrics it is still not fully clear what exactly these metrics are measuring and how to face challenges that arise from their usage. Hence, it is necessary to further evaluate altmetrics.

Evaluating altmetrics

"[T]he act of citing has been an essential part of scholarly communication in modern science, whereas researchers are still exploring how to use social media" (Haustein 2016, p. 414). Considering this, it is not always clear why a scientific publication is mentioned on Twitter, Facebook or scientific blogs. Not necessarily is credit given to a publication via a tweet or post. Social media and blogs are mediums on which it is easy to criticize everything, including publications (Shema et al. 2012). In addition, some tweets are not even composed by humans, but by bots automatically mentioning a paper (Haustein et al. 2016).

Based on the main user group writing posts and mentioning publications, the impact of altmetrics can assume different shapes from educational to societal impact (Sud and Thelwall 2014). Likewise, each metric may measure a different type of impact. A tweet only allowing 140 characters may have a different focus than a blog post of any length or the act of saving a document. So far, only a few studies concentrate on qualitative investigations of altmetrics and their actual meaning. In order to better understand the different types of metrics, Haustein et al. (2015) present a framework of types of acts referring to research objects. The authors differentiate between scholarly documents and agents (e.g., a researcher) as research objects. In this work, the focus is set on documents, hence only this part of the framework will be considered. Haustein et al. (2015) further discuss citation and social theories in order to better understand why certain events occur. They conclude that these theories cannot fully explain the access, appraisal and application of a research object and ask for more qualitative investigations like content analyses and user surveys.

Still, it is discussed if altmetrics serve as an indicator to evaluate scientists, journals and institutions. Also publishers might have an interest in following the development of the metrics of their publications. Hence, "given that there are many different parts of the social web, evidence about the value and relative importance of each one for altmetrics would be useful for publishers seeking to use them effectively" (Sud and Thelwall 2014, p. 1132). In addition, in the Leiden Manifesto we read "Research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise. But this quantitative information must not be allowed to morph from an instrument into the goal" (Hicks et al. 2015).

Haustein (2016) describes three major challenges in altmetrics: heterogeneity, data quality and dependencies. Considering heterogeneity, the author emphasizes that a lot of different types of altmetrics exist, which makes it difficult to clearly define the term. Further, it is concluded that more quantitative and qualitative research is necessary to understand and evaluate altmetrics. A lack of data quality is another important critical factor that is mentioned. Finally, dependencies exist on technical conditions like the provision of an API (Haustein 2016).

Regarding these different types of challenges, also Sud and Thelwall (2014, p. 1140) emphasize that various evaluation methods—like calculating correlations, conducting creator interviews and content analysis—are necessary and "can all give evidence about the value or meaning of altmetrics". Also Jobmann et al. (2014) conclude that coverage and metric counts can vary across different data aggregators and call for more analyses in comparing altmetric providers in order to detect challenges in altmetric data collection. Similarly Zahedi et al. (2014) compare the metrics of 1000 PLOS publications from three altmetric providers. They detect differences in the number of Mendeley readers, tweets and Facebook counts and articulate that the consistency of altmetric providers has to be improved. This study aims at further analyzing two altmetric aggregators in terms of the consistency of their metrics. Subsequently, the methods for this investigation are brought forward.

Methods

Sud and Thelwall (2014) recommend to organize the evaluation of altmetrics in a logical order. First, simple quantitative tests should be conducted, as collecting and evaluating data performs a transparent process. In particular, the authors refer to the calculation of correlations between altmetrics and citations. In a second stage, content analysis of selected data sources should be applied in order to categorize contexts and user motivations. At last, interviews and pragmatic evaluations for information retrieval systems should be conducted for more detailed analyses, but are also time consuming (Sud and Thelwall 2014).

Following these recommendations, this work will cover the first step mentioned by Sud and Thelwall (2014). Therefore, a random set of 5000 journal articles has been chosen and used for the analysis of the metrics. The quantitative investigations include the coverage of different metrics across two aggregators and the correlations between these. In contrast, correlations between citation counts and altmetrics will not be considered in this investigation, as there are already several studies, all with similar results (Erdt et al. 2016).

In order to obtain a testing set of articles that can be used to compare the metrics of several altmetric data aggregators, the professional database Web of Science has been consulted. This provider has been chosen as it represents a multidisciplinary platform that can be used for bibliometric analyses (Moed 2009). Altmetrics claim to be fast indicators for research assessment, for this reason relatively recent publications have been chosen for this investigation. Hence, a multidisciplinary set of journal articles with publication year 2015 and publication language English has been retrieved resulting in a set of 1,499,100 documents on September 27th, 2016. From this huge amount a random set of 5000 articles has been chosen. One condition that had to be fulfilled for choosing the publications was the presence of a DOI, as this can be used to identify a publication unambiguously and retrieve altmetrics in the different aggregators.

The next step was the selection of altmetric providers for further analyses. In this study, PlumX and Altmetric.com have been chosen as they are not limited to publications of one publisher (like PLOS article level metrics) and thus allow an evaluation of a more general set of articles. Altmetric.com provides interested users the possibility to use an open API.[1] With the help of the API and a python script it was possible to retrieve all the metrics from Altmetric.com via DOI, including the Altmetric Attention Score. PlumX does not offer an API that can be used freely by individuals. Thus, the Plum Analytics team has been contacted in order to get access to the metrics. As a result a personal site has been created by the Plum Analytics staff, that lists all the metrics from PlumX and allows to download all the data in a CSV file.

For evaluating the metrics of PlumX and Altmetric.com, descriptive statistics have been used as a first approach to get an impression of the coverage of the publications in the different sources. In addition inter- and intra-correlations between the metrics and altmetric providers have been calculated by using Spearman's rank correlation (Spearman 1904). The advantage of this type of correlation is that it fits well with altmetrics data as there are a lot of zeros and upward outliers and thus a normal distribution cannot be assumed. The correlations have been calculated by using the *R* functions `cor`. Thereby, intra-correlation for the metrics within one aggregator can be calculated pairwise on the basis of all the DOIs recognized by the service and the corresponding metric counts. Inter-correlations for those metrics that are covered by both aggregators are similarly calculated. Thereby the intersection of articles captured by PlumX and Altmetric.com has been used.

Results

In the following section, results regarding the coverage of the two aggregators and their metrics are presented. After that, the findings on intra- and inter-correlations are depicted.

[1] http://api.altmetric.com/.

Coverage

From the 5000 DOIs that have been collected from the Web of Science, 4936 (99%) could be traced by PlumX. The number of DOIs having hits on Altmetric.com in contrast sums up to 1955 (39%). Here the aggregators' approaches play an important role: while Altmetric.com aims at providing metrics that are accessible for everyone, PlumX creates customized products. This should be taken into account when comparing the coverage of the aggregators. The distribution of the DOIs on the disciplines is depicted in Fig. 1. Most DOIs from the initial set have been marked as Multidisciplinary (1707), which represents 34% of all articles. As well, the Natural Sciences are represented by a large set of 1425 articles (29%). Computer Science, Engineering and Mathematics has been dedicated to 695 articles (14%). With similar values, the Life Sciences (9%), Medicine and Health Sciences (7%) and Arts, Humanities and Social Sciences (7%) make their contribution.

PlumX provides metrics for 98 to 99% of the articles from the individual disciplines. Overall, with 4936 covered articles, PlumX captures at least one event for 99% of the investigated articles. Altmetric.com in contrast shows more variations in the coverage of different disciplines. The aggregator only covers 15% of the articles in Computer Science, Engineering and Mathematics, 29% of the Natural Sciences and 38% of the multidisciplinary articles. Higher are the values for Medicine and Health Sciences (57%), Arts, Humanities and Social Sciences (67%) and Life Sciences (75%).

There are some data sources and metrics that are covered by only one aggregator and also sources that are considered by both of them. Mendeley[2] readers, tweets on Twitter,[3] Facebook[4] shares, comments on Reddit,[5] links on Wikipedia,[6] Google+[7] mentions as well as news and blog mentions are events that Altmetric.com and PlumX take into account. Altmetric (2017) provides detailed information on the start and end date of the coverage of all its data sources. Most of the sources are being tracked since October 2011 when Altmetric.com started its service. Among them are Twitter, Facebook, news, blogs, Reddit, Mendeley and CiteULike whereas for the latter coverage ended in December 2014.

Figure 2 depicts the number of articles that are covered by the aggregators. To make proper comparisons, metrics covered by only one aggregator are not represented here. It becomes apparent that Mendeley is by far the most covered data source for PlumX, but also Altmetric.com. From the initial set, 4566 publications (91%) are recognized as having at least one Mendeley reader according to PlumX. Altmetric.com in contrast only observes 1899 articles with Mendeley readers which accounts for 38% of the initial set.

Beside Mendeley, Twitter is the data source covered by the aggregators second most frequently. In this case, it is conspicuous that the highest number of articles is covered by Altmetric.com. Even 1789 DOIs (36%) have been mentioned in tweets according to the aggregator. PlumX in contrast only recognizes 1031 DOIs (21%) mentioned in tweets. Except for Wikipedia links, the coverage for the other metrics are similarly higher for Altmetric.com.

[2] https://www.mendeley.com/.

[3] https://www.twitter.com/.

[4] https://www.facebook.com/.

[5] https://www.reddit.com/.

[6] https://www.wikipedia.org/.

[7] https://plus.google.com/.

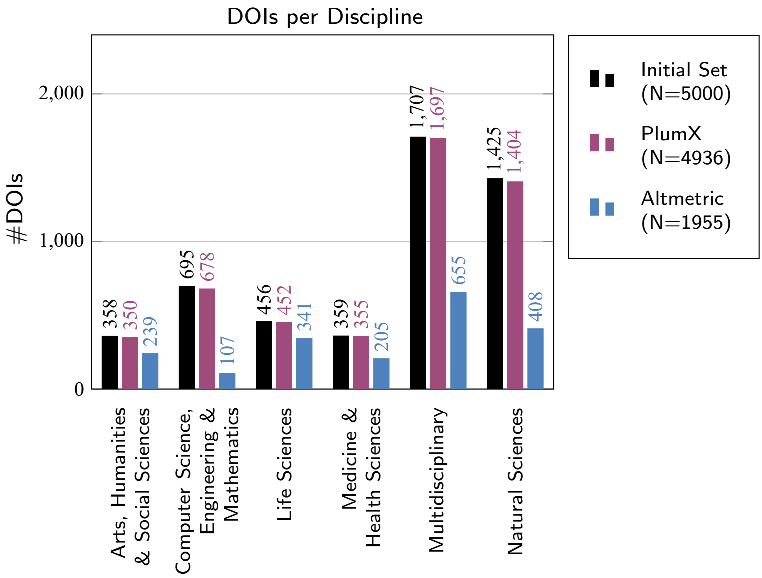

Fig. 1 Number of DOIs in initial set and captured by the aggregators per discipline

Intra- and inter-correlations of altmetrics

It is not clear what all the different metrics are actually measuring and which metrics are measuring similar kinds of impact or attention. Hence, it will be investigated how the individual metrics correlate with each other within one aggregator (intra-correlations). Further, the aggregators seem to notice and gather different counts for the same data sources which has become apparent in the above section. The inconsistencies of the two aggregators will hence be analyzed in terms of inter-correlations.

Intra-correlations have been calculated for the shared metrics provided by the aggregators. For the altmetric provider PlumX, intra-correlations are listed in Table 2. Most correlations are considerably weak, which shows that attention as measured by altmetrics cannot be generalized.

Within social media metrics, the highest correlation values are produced by Twitter and Facebook ($r = 0.35$). Google+ correlates with Facebook ($r = 0.22$) and Twitter ($r = 0.14$) on a low level. News and blogs do not relate to each other ($r = 0.07$). Wikipedia mentions do not correlate with any of the other metrics.

For the second aggregator—Altmetric.com—correlations could be calculated supplementary for the Altmetric Attention Score (Table 3). For the data sources Reddit and Wikipedia no correlation could be determined. In contrast to PlumX the data of Altmetric.com reveals a moderate correlation of news and blogs ($r = 0.43$). Again, it should be examined how the aggregators gather their data.

Further, for social media metrics, the correlations within are lower than calculated for PlumX and range from 0.15 (Facebook–Google+) to 0.19 (Facebook–Twitter). Regarding Facebook, it should again be considered that only shares are included on the part of Altmetric.com. In comparison to other metrics, Mendeley correlates the most with Twitter, having a similar value as determined for PlumX ($r = 0.32$).

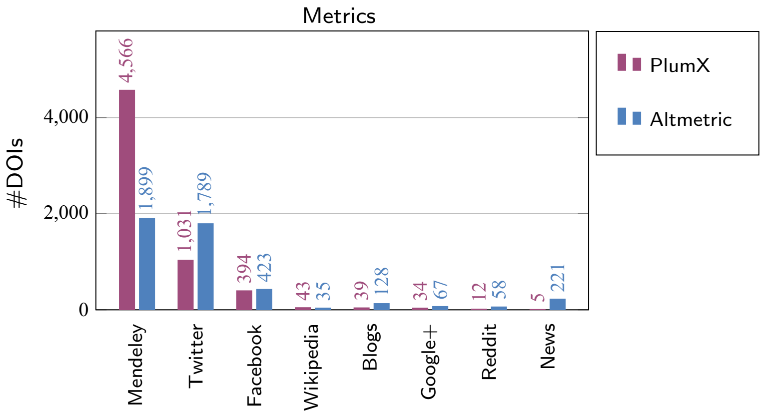

Fig. 2 Number of DOIs captured by the aggregators for the different metrics ($N = 5000$)

Table 2 Correlations of the metrics provided by PlumX

	Facebook	Blogs	Google+	News	Reddit	Twitter	Wikipedia
Facebook							
Blogs	0.20*						
Google+	0.22*	0.19*					
News	0.02	0.07*	0.07*				
Reddit	0.19*	0.15*	0.15*	0.06*			
Twitter	0.35*	0.15*	0.14*	0.02	0.15*		
Wikipedia	0.04***	0.04***	0.05***	0.07*	0.05**	0.05**	
Mendeley	0.18*	0.11*	0.09*	0.02	0.07*	0.27*	0.07*

*$p < 0.0001$; **$p < 0.001$; ***$p < 0.01$

Table 3 Correlations of the metrics provided by Altmetric.com

	Facebook	Blogs	Google+	News	Reddit	Twitter	Wikipedia	Mendeley
Facebook								
Blogs	0.14*							
Google+	0.15*	0.23*						
News	0.21*	0.43*	0.20*					
Reddit	0.09**	0.15*	0.10*	0.12*				
Twitter	0.19*	0.21*	0.17*	0.17*	0.09*			
Wikipedia	0.02	0.09*	0.06***	0.04	0.04****	− 0.01		
Mendeley	0.17*	0.18*	0.08**	0.20*	0.06****	0.32*	0.05****	
Attention Score	0.26*	0.40*	0.22*	0.54*	0.11*	0.70*	0.15*	0.38*

*$p < 0.0001$; **$p < 0.001$; ***$p < 0.01$; ****$p < 0.05$

Table 4 Inter-correlations of the metrics provided by the aggregators

	Facebook	Blogs	Google+	News	Reddit	Twitter	Wikipedia	Mendeley
PlumX–Altmetric	0.29*	0.46*	0.28*	0.11*	0.41*	0.49*	0.82*	0.97*

*$p < 0.0001$

The Altmetric Attention Score should be considered separately. Here, relatively high correlations have been calculated, as most data sources are considered in the calculation of the score. The highest value has been identified for Twitter ($r = 0.7$). This is followed by news ($r = 0.54$), blogs ($r = 0.4$), Mendeley ($r = 0.38$) and Facebook ($r = 0.26$). These sources do contribute to the calculation of the Altmetric Attention Score with different weightings. Although Twitter is only considered with a weight of $w = 1$, it attains by far the highest correlation value. News and blogs are dedicated a weight of $w = 8$ and $w = 5$ respectively. Mendeley in contrast is not included in the calculation of the score, but still achieves a moderate value. Lastly, Facebook only gets a weight of $w = 0.25$.

In comparing the correlations of several metrics within the aggregators, it has already become apparent, that there are variations in-between the altmetric providers. In the following section it will be further analyzed how the metrics of the two aggregators correlate with each other.

In Table 4 the correlation values are depicted. Overall PlumX and Altmetric.com still achieve a moderate correlation value of $r = 0.50$. Considering the concrete metrics of the aggregators, the most striking value is achieved for Mendeley readers. Here, the aggregators correlate on a high level of $r = 0.97$. Similarly, Wikipedia achieves a good value of $r = 0.82$. Google+ produces a low value of $r = 0.28$ and Reddit a moderate correlation of $r = 0.41$. Tweets also show a medium correlation of $r = 0.49$. For Facebook, the value only amounts to $r = 0.29$, which can be explained by the circumstance that likes and comments are not considered by Altmetric.com but by PlumX.

For blogs, a moderate value of $r = 0.46$ has been calculated. The lowest value is achieved with regard to news mentions ($r = 0.11$). As the aggregators cover a different set of blogs and news, this is not surprising. Altmetric e.g., tracks more than 9000 academic and non-academic blogs (Altmetric 2015). Similarly, PlumX tracks blogs from a manually curated list and distinguishes between blogs in general and blogs related to economics (Plum Analytics 2016). Regarding news, Altmetric picks mentions from RSS feeds from news websites and also includes non-English news outlets (Altmetric 2016b). PlumX covers about 55,000 news sources all over the world whereby three main sources for blogs and news are mentioned: mass media sources from Newsflo[8] (a service of Elsevier), ACI's index of academic blog content[9] and a manually curated list for PlumX of blogs and news sources (Allen 2017).

8 https://www.elsevier.com/solutions/newsflo.

9 https://aci.info/scholarly-blogs/.

Discussion and conclusion

Based on the necessity for more analyses concerning the consistency of different altmetrics, 5000 articles covering a wide range of disciplines have been retrieved and analyzed concerning their metrics across the data aggregators PlumX and Altmetric.com. The investigations included analyses on coverage across the altmetric providers, as well as intra- and inter-correlations of the metrics.

The aggregator PlumX provides customized products and covers a majority of the articles in the initial data set. In contrast, Altmetric.com does not provide metrics for even half of the investigated documents. The aggregators gather data from a different set of altmetric sources. They agree on the admission of data from Twitter, Facebook, Mendeley, Google+, Reddit, blogs, news and Wikipedia, but show variations for example, considering the admission of blogs and news sources. The selection decision illustrates that the popularity of altmetrics differs depending on the source but also on the time. The example of Altmetric.com showed that some attention sources as CiteULike are no longer tracked by the aggregator which makes altmetrics a fast-moving research subject. Hence, it should be further explored if the sources chosen in common like Twitter, Facebook and blogs in fact are an appropriate tool for assessing research.

Considering the common attention sources, the aggregators show variations in the coverage of the data set. Both altmetric providers agree on a high coverage of Mendeley considering the 5000 initial articles, though the numbers for PlumX are even double as high as for Altmetric.com. Studying however the social media data sources Twitter and Facebook, the coverage of Altmetric.com is higher than PlumX's. These discrepancies indicate that the data gathering process of the aggregators seems to differ immensely. This makes altmetrics vulnerable for manipulations and incorrect conclusions.

There are also various concepts on measuring impact among the aggregators which is expressed in the differences of counted objects e.g., regarding Facebook. When different events can be counted—e.g., shares, likes and comments—the aggregators themselves have to make a decision what to track. In some cases the concrete events counted by the aggregators are not explained or at least hard to find. For a user it may be irritating reading a certain number without knowing what it actually means.

The differences between individual metrics as well as the aggregators come even more clear by considering intra- and inter-correlation. The correlations between different social networks are low to moderate, for Twitter and Facebook a value of 0.35 could be achieved. Accordingly, it is questionable if these should be considered as comparable and classified as the same type as it is done in altmetric frameworks and also by the aggregators. Considering Altmetric.com, again only low to medium correlations have been calculated. In contrast, the Altmetric Attention Score correlates positively with Twitter, although the weighting of tweets is relatively low regarding the calculation of the score. Mendeley correlates positively with the Altmetric Attention Score with a value of 0.38. Nonetheless it is not included in calculating the score, although it has stuck out in many investigations as it has the highest correlations with classical citation counts. Mendeley further covers more publications than the other metrics and thus seems to be well accepted by a wide range of users. For the score, it should also be scrutinized why Weibo has a higher weight than Facebook, even though it is no longer tracked by the aggregator. Agreeing with Mukherjee et al. (2018), the Altmetric "donut" should be called into question. It seems to be a nice feature for a user to have a visual representation of the metrics but its calculation is questionable. Without consideration of the Altmetric Attention Score, the highest correlations have been calculated between news and blogs with a moderate value of 0.43. According to the metric counts of PlumX a correlation between these sources is virtually not existing with a value of 0.07. This again demonstrates that the aggregators select different news and blog sources, which should be taken into account when relying on one of the altmetric providers for assessing research.

Beside the relations between the metrics of one aggregator, the overall correlations within the altmetric providers have been calculated. Overall, Altmetric.com and PlumX correlate on a medium level. Low values have thereby occurred for news and Reddit. A metric, for which the highest consistency in terms of inter-correlations has been achieved is Mendeley. Again, this data source sticks out in the evaluation. Regarding coverage, correlations with citations as well as inter-correlations between the aggregators Mendeley comes off well. Accordingly, this could be a promising metric regarding the evaluation of science. However the comparison with citations presumes that this classical metric is an adequate measure for assessing research. Therefore, more analyses should be conducted on users' motivations for working with the reference manager. As no written texts are available, content analysis is not the appropriate method for this source, but interviews and questionnaires could work well.

There is currently a large number of altmetric sources which have to be considered in differentiated ways. A common standard or framework is still lacking. The possibility to have a fast indicator to measure attention is an advantage of altmetrics. Though, the coverage differs among the data sources. In order to evaluate a large amount of publications efficiently, it is necessary to have comparable data for a majority of the research subjects. At the moment, this holds true only for Mendeley readers and to some extent tweets and Facebook posts. The huge amount of metrics further leads to different assessments of publications. This leads to a lack in efficiency as it is not clear which measure and aggregator to rely on. It should be further explored what additional factors influence the inconsistencies of altmetrics. Parameters like the publication year or subject area might have an impact on the examined discrepancies. Hence, further studies should extend the analysis in this regard.

The results also show that altmetrics are too complex to aggregate them to one measure. Nonetheless, the aggregators facilitate the overview of the different metrics and should help in assessing altmetrics. Variations in the selected sources questions this assessment. For sources like blogs where selections on the coverage have to be made, the providers play an important role.

In conclusion, altmetrics can be consulted in order to obtain an impression on the attention a research object has received. Thereby, attention cannot be generalized as different altmetrics sources lead to different assessments of a research object. Further, instead of merely counting everything that is possible, some suitable data sources should be selected. Further analysis is needed to assess the sources also in a qualitative way. Here, the aggregators' feature to access the concrete contents should be given priority as it is making research more transparent and goes beyond mere counting. Altmetrics hence have a lot of potential but also need to be applied correctly. What 'correctly' means in this context has to be further explored in qualitative studies.

References

Allen, P. (2017). World-wide news coverage in PlumX. https://plumanalytics.com/world-wide-news-coverage-plumx/. Accessed 2 Dec 2017.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of Documentation*, *53*(4), 404–426. https://doi.org/10.1108/EUM0000000007205.

Altmetric. (2015). Sources of attention. https://www.altmetric.com/about-our-data/our-sources/. Accessed 30 Jan 2018.

Altmetric. (2016a). How is the altmetric attention score calculated? https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-. Accessed 30 Jan 2018.

Altmetric. (2016b). Mainstream media outlets. https://www.altmetric.com/about-our-data/our-sources/news/. Accessed 2 Dec 2017.

Altmetric. (2017). When did altmetric start tracking attention to each attention source? https://help.altmetric.com/support/solutions/articles/6000136884-when-did-altmetric-start-tracking-attention-to-each-attention-source-. Accessed 30 Jan 2018.

Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, *103*(3), 1123–1144. https://doi.org/10.1007/s11192-015-1565-y.

Erdt, M., Nagarajan, A., Sin, S.-C. J., & Theng, Y.-L. (2016). Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, *109*(2), 1117–1166. https://doi.org/10.1007/s11192-016-2077-0.

Garfield, E. (1955). Citation indexes for science. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.123.3185.61-a.

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

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 and scholarly communication* (pp. 372–405). Berlin/Boston: De Gruyter.

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. https://doi.org/10.1002/asi.23456.

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

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, *5*(3), 446–457. https://doi.org/10.1016/j.joi.2011.04.002.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

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.

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.

Lohr, C. (2017). New in captures: CiteULike readership. https://plumanalytics.com/new-in-captures-citeulike-readership/. Accessed 12 Dec 2017.

Martin, J. (2015). Responsible metrics. In J. Wilsdon (Ed.), *The metric tide. Report of the independent review of the role of metrics in research assessment and management* (pp. 134–147). HEFCE. https://doi.org/10.13140/RG.2.1.4929.1363.

Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, *79*(4), 606–623. https://doi.org/10.1086/354848.

Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum Immunologiae et Therapiae Experimentalis*, *57*(1), 13–18. https://doi.org/10.1007/s00005-009-0001-5.

Mukherjee, B., Subotić, S., & Chaubey, A. K. (2018). And now for something completely different: The congruence of the altmetric attention score's structure between different article groups. *Scientometrics*, *114*(1), 253–275. https://doi.org/10.1007/s11192-017-2559-8.

Nicholas, D., & Rowlands, I. (2011). Social media use in the research workflow. *Information Services and Use*, *31*(1–2), 61–83. https://doi.org/10.3233/ISU-2011-0623.

Plum Analytics. (2016). Mention metrics. https://plumanalytics.com/learn/about-metrics/mention-metrics/. Accessed 2 Dec 2017.

Plum Analytics. (2017). Leadership. http://plumanalytics.com/about/leadership/. Accessed 30 Jan 2018.

Priem, J. (2010). I like the term #articlelevelmetrics, but it fails to imply *diversity* of measures. Lately, I'm liking #altmetrics. https://twitter.com/jasonpriem/status/25844968813. Accessed 30 Jan 2018.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. http://altmetrics.org/manifesto. Accessed 30 Jan 2018.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PLoS ONE*, *7*(5), 1–8. https://doi.org/10.1371/journal.pone.0035869.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159.

Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, *98*(2), 1131–1143. https://doi.org/10.1007/s11192-013-1117-2.

Thelwall, M. (2016). Webometrics and altmetrics: Home birth vs. hospital birth. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication* (pp. 337–346). Berlin/Boston: De Gruyter.

Torres-Salinas, D., Robinson-Garcia, N., & Jiménez-Contreras, E. (2016). Can we use altmetrics at the institutional level? A case study analysing the coverage by research areas of four Spanish Universities. In *Proceedings of the 21st international conference on science and technology indicators*, València, Spain. http://dx.doi.org/10.4995/STI2016.2016.xxxx.

Weller, K. (2015). Social media and altmetrics: An overview of current alternative approaches to measuring scholarly impact. In I. M. Welpe, J. Wollersheim, S. Ringelhan & M. Osterloh (Eds.), *Incentives and performance: governance of research organizations* (pp. 261–276). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-09785-5_16.

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

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. altmetrics 14. In *Workshop at the Web Science Conference Bloomington*, USA, pp. 4–6. https://doi.org/10.6084/m9.figshare.1041821.