The coverage of blogs and news in the three major altmetric data providers

José Luis Ortega[1]

[1] *jortega@iesa.csic.es*
Institute for Advanced Social Sciences, CSIC, Campo Santo de los Mártires, 7 14004 Córdoba (Spain)

Abstract
The objective of this study is to compare the coverage of the three main altmetric data providers (PlumX, Altmetric.com and Crossref Event Data) according to blog posts and news. The paper emphasizes the coverage the events that mention publications with the purpose of observing the media coverage and overlap between services. More than 100,000 random publications from Crossref were searched in the three providers. The link, title and source of the events that mention each document were retrieved. Results show that the number of publications mentioned in blogs (9.9%) and news (4.8%) is very low, being Altmetric.com the service with more documents mentioned by blogs (6%) and news (3.7%). Altmetric.com also covers more blogs (37.8%), while PlumX collects a large number of news media (36.5%). The overlap between the altmetric providers is rather low in publications (7.4%), events (0.5%) and sources (4.1%). The study concludes that the employment of several providers is necessary to undertake any reliable altmetric study.

Introduction
Altmetric data providers have gained great importance in the scholarly publication system because they capture and make public the mentions that receive academic documents in different web environments. Exploring an endless number of media, social networks, blogs, reference managers and web databases, these services find and index events that mention research outputs. This information is acquiring great value for different stakeholders. Publishers esteem altmetric data because they can track the usage and valuation of the papers that they publish, observing who their audiences are and how they respond to their publications. Policy makers and funding agencies can appreciate the impact of the research that they fund in different social environments.

However, bibliometricians and altmetricians are the most demanding users because they utilize these tools as research instruments to analyse the meaning and behaviour of these metrics and to observe their involvement in evaluation activities. Altmetric studies are increasingly depending on these services to carry out their analyses. This dependence is based on the technical difficulty of tracking mentions from whole the Web and the need to establish agreements with third parties (Twitter, Facebook, etc.). This situation makes very difficult for researcher to undertake altmetric studies by their own means.

This fact is causing growing concern about the reliability and accuracy of these platforms as data providers because the validity of the results is strongly determined by the source used for the analysis. Several studies have addressed this issue, comparing the counts supplied by each platform in a sample of publications (Meschede and Siebenlist, 2018; Zahedi and Costas, 2018; Ortega, 2018). Results, in each case, have evidenced important differences in coverage of publications and metric counts. These results have been criticized because the comparison of total counts does not allow to study the overlap between providers (Wass, 2018). That is, comparing the total number of mentions in different platforms does not permit to know how many of them are repeated. Then, a fair comparison should be to identify all the posts that mention a publication and check if they are the same or not across different platforms. However, it is not easy to obtain the original documents (tweets, blog posts, news, etc.) that mention a publication from these providers, which supposes an important problem to carry out studies about the overlap among altmetric providers.

This study tries to fill this gap, analysing the largest sample of publications (100,000 documents) ever used for comparing the three major altmetric providers (Altmetric.com, PlumX and Crossref Event Data (CED)). In this case, the study is limited to compare the coverage of blog posts and news with the aim of benchmarking not only the overlap between providers but also of checking the size of their lists of blogs and news outlets.

Literature review

In spite of the growing importance of data providers in the altmetric studies, the literature about the functionalities and working of these platforms is not very large. Adie and Roe (2013) were the first ones who detailed how Altmetric.com tracks the mention of papers on the Web. Trueger et al. (2015) made a critical review of the Altmetric Score, although, it was Gumpenberger et al. (2016) who expressed the strongest criticism about this indicator. In the case of PlumX, Champieux (2015) and Lindsay (2016) described the utilities of the service, while Wong and Vital (2017) analysed the implementation of the tool in a specific organization. However, no studies have been yet published about the functionality of CED mainly due to this product is still in beta.

Nevertheless, many other studies have analyzed the coverage of these services, describing the proportion of altmetric events in different samples. Thelwall et al. (2013) performed the earliest distribution of metrics in Altmetric.com, finding a greater proportion of papers mentioned in Twitter and Facebook. Robinson-García et al. (2014) also analyzed the coverage of this provider and they found that 87.1% of articles had at least one tweet and 64.8% one Mendeley reader. In a similar way, Bornmann (2014) explored a set of articles from Altmetric.com and he observed that 71% of articles were tweeted and a moderated proportion of documents were mentioned in Facebook (31%). More recently, Thelwall (2018) studied the coverage of Social Sciences, Arts and Humanities found a low prevalence (less than 12%, excepting tweets). According to PlumX, it worth mentioning the work of Torres-Salinas et al. (2017) about the collection of books. Their results showed that the distribution of events for books is rather different than for articles. Ortega (2018a) used PlumX data to track the life cycle of several altmetrics, observing that the most frequent ones were Mendeley readers and Tweets.

Specifically, several studies have focused on the blogs posts as altmetric indicators. Fausto et al. (2012) were the first ones to explore the relationship between blog posts and citations. Their result shown a positive correlation between post views and citations. Shema et al. (2014) found that articles receiving blog mentions close to their publication date receive more journal citations. These same authors (Shema et al., 2015) found that reviews and multidisciplinary top-tier journal articles were overrepresented in blog mentions. Jamali and Alimohammadi (2015) observed that discussion and criticism were the two main categories of motivations for citing articles in blogs. Contrarily, literature about the mention of research articles in news outlets is almost non-existent. Only, it worth to mention studies that describe the proportion of articles cited in news. According to Altmetric.com, Bornmann (2014) observed a moderated proportion of documents mentioned in news (13%), while Fraumann et al. (2015) found an important bias towards U. S. sites. MacLaughlin et al. (2018) studied the features that improve the popularity of research articles mentioned in news.

However, more papers have performed comparative studies between altmetric aggregators. Jobmann et al. (2014) were the first ones to compare the coverage and counts of ImpactStory, Altmetric Explorer, Plum Analytics and Webometric Analyst by research areas. Their results showed important divergences between services, being Plum Analytics the platform that better covered Mendeley and Facebook data, while Altmetric.com highlighted gathering blogs, news and CiteULike data. Zahedi et al. (2015) explored the consistency of data across Altmetric.com, Mendeley and Lagotto. They also detected significant differences, finding that Altmetric.com gathered more tweets, but it was less accurate collecting Mendeley readers. Baessa et al. (2015)

evaluated several altmetric services for their institutional repository and they recognized that Altmetric.com had a better coverage of blogs, news and government documents, while PlumX was most exhaustive covering different formats as books or reports. Kraker et al. (2015) studied the gathering of research data in Figshare, PlumX and ImpactStory. They observed that PlumX detected considerably more items in social media and higher scores than ImpactStory. Peters et al. (2016) extended their former study (Peters et al., 2015) with the inclusion of Altmetric.com. Their results confirmed that PlumX was the best provider for covering non published materials such as research data. Meschede and Siebenlist (2018) compared Altmetric.com and PlumX, finding that less than half of the publications analyzed were included in Altmetric.com, while PlumX covered almost the totality (99%). Zahedi and Costas (2018) performed the most exhaustive comparison between data providers, finding substantial differences in the metrics offered by these platforms. Bar-Ilan et al. (2018) compared Mendeley, Altmetric.com and PlumX in two time spots (2017 and 2018). Their results showed that the overlap between PlumX and Altmetric.com increased. Torres-Salinas et al. (2018) compared the coverage of books in Altmetric.com and PlumX and they concluded that they are rather complementary than comparable tools. Finally, Ortega (2018b) benchmarked the metrics counts of Altmetric.com, PlumX and CED, observing that Altmetric.com is the best aggregator of blog posts, news and tweets; PlumX of Mendeley readers; and CED of Wikipedia citations.

Objectives

The main objective of this study is to contrast, for first time, the coverage of the three main altmetric data providers according to blog posts and news. The aim is not only comparing the coverage of publications and number of counts, but also exploring the own events that mention the publications with the purpose of observing the media coverage and overlap between services. In particular, the following research questions were addressed:

- To what extent do the three providers cover the publications of the sample and what is the overlap between them?
- To what extent do the three providers capture blog posts and news that mention the publications of the sample and what is the overlap between them?
- What is the media coverage of the three providers and what is the overlap between them?

Methods

This study is particularly focused on blogs and news mentions. The reason for analysing only these metrics is that they do not come from a specific platform such as tweets (Twitter) or readers (Mendeley), but from a large number of media. This fact makes possible comparing the coverage of both documents and events, and the sources of those events. Another reason is that blogs and news are comparable metrics in the three providers.

Altmetric providers

PlumX: PlumX (plu.mx/plum/g/samples) is a provider of alternative metrics created in 2012 by Andrea Michalek and Michael Buschman from Plum Analytics. This product is addressed to the institutional market, offering altmetric counts of publications for particular institutions. PlumX is the aggregator that offers more metrics, including citation and usage metrics (i.e. Views and Downloads). It covers more than 52.6 million of artifacts, being the largest altmetric aggregator (Plum Analytics, 2018). In 2017, Plum Analytics was acquired by Elsevier (www.elsevier.com), tracking now the online presence of any article indexed in Scopus database (Elsevier, 2017). This agreement with Elsevier also caused that PlumX used Newsflo (an Elsevier company) as news data provider (Allen, 2017). On the other hand, PlumX also

includes blog mentions from ACI Scholarly Blog Index (ACI, 2016). However, PlumX does not provide any information about the number of blogs and media covered.

Altmetric.com: It was the first altmetric provider and it was initiated in 2011 by Euan Adie, with the support of Digital Science (www.altmetric.com). Unlike PlumX, Altmetric.com is centered in the publishing world, signing agreements with publisher houses to monitor the altmetric impact of their publications. This information is accessible through a public API. Today, Altmetric.com tracks the social impact of close to 9 million of research papers (Altmetric.com, 2018). Altmetric.com monitors around 11,000 blogs (Altmetric.com, 2018), although it does not make available the list of sources. According to news, Altmetric.com collects a list of 1,300 news outlets, which could be expanded to 80,000 thanks to a partnership with Moreover.com (A LexisNexis Company) (Williams, 2015). This list is publicly available on the site.

Crossref Event data (CED): CED is the youngest service. Created in 2016, it is still in beta (www.crossref.org/services/event-data). Unlike the previous ones, CED is not a commercial site and it provides free access to their data through a public API. Another important difference is that it does not provide metrics, but it only displays information about each altmetric event linked to a DOI identifier. For instance, it shows the information about the mention of an article on Twitter (date, user, tweet, etc.), but it does not show a count of the number of tweets. For that reason, CED's data would have to be processed to be comparable with the other services. CED does not distinguish between blogs and news. It defines three categories, *wordpressdotcom*, *web* and *newsfeed*s to group blogs and news. In addition, category *reddit-links* includes links from Reddit that point external sources such as blogs and news. A manual inspection evidenced that the media and blogs are equally classified as *web* or as *newsfeed*, and sometimes in both categories at the same time. Due to this, the distinction between blog and news is based on the matching with the other data providers. In the case of mentions that do not match, then a manual classification was done.

Data extraction

This study aims to compare the coverage of blogs and news mentions by the three major altmetric providers. A random sample of 100,529 DOIs from Crossref were extracted to detect the number of publications covered by these aggregators. These publications were obtained from Crossref API with the conditions of being journal articles and published from 2012 (https://api.crossref.org/works?sample=100&filter=type:journal-article,from-pub-date:2012-01-01). 2012 was elected because is the time window sufficiently broad to capture the impact of the sample in blogs and news.

Next, this list was searched in the three data providers. In the case of Almetric.com, Altmetric ID was obtained from the Altmetric API (api.altmetric.com/v1/doi/), and then it was used to extract data about blogs and news directly from the web site (www.altmetric.com/details/). This is due to the API only shows counts and not the links and content of these mentions. In the case of PlumX, DOIs were searched in the web site of PlumX (plu.mx/plum/a/?doi=). Finally, information of CED was extracted from the API (query.eventdata.crossref.org/events?filter=obj-id:). In the three cases, several SQL scripts were written to scrape the data from websites and APIs. This process was performed during the second fortnight of August 2018.

Results

Publications

Our first objective is to know the overlap of blogs and news between the three data providers. Table 1 shows the number and proportion of articles indexed in these providers, the number

and proportion of articles with any mention, and the relationship between blogs and news in each platform.

Table 1. Distribution of articles in the three major data providers

	Altmetric.com	*PlumX*	*CED*	*Sample*
Articles	34,583	99,697	100,116	100,529
% Sample	34.4%	99.2%	99.6%	11,728
Mentioned articles	7,457	7,274	1,698	11,7%
% Sample	7.4%	7.2%	1.7%	
% Provider	21.6%	7.3%	1.7%	9,990
Blogs	5,995	5,741	1,222	11,728
% Sample	6.0%	5.7%	1,222	9,990
% Provider	17.3%	5.8%	1.2%	9.9%
News	3,712	2,784	1.2%	
% Sample	3.7%	2.8%	662	4,862
% Provider	10.7%	2.8%	0.7%	4.8%

CED is the service that indexes more journal articles (99.6%) because Crossref provides its database of articles to CED. The remaining .4% could be to technical errors during the data extraction process. PlumX is the second one in coverage with 99.2% of the sample, which it makes clear its strong capability to gather publications. Altmetric.com, however, only covers the 34% of the sample, which confirms the reduced size of this tool (Meschede and Siebenlist, 2018; Ortega, 2018). Nevertheless, not all the publications indexed in a provider have mentions. Concretely, regarding to blogs and news mentions, Altmetric.com is the service that have more articles mentioned in blogs and news (7.4%), followed shortly by PlumX (7.2%) and CED (1.7%). These results show that PlumX and Altmetric.com, independently of their sizes, gather a similar proportion of articles discussed in media. Other important result is that CED, in spite of its important coverage of publications, captures only a small fraction of events. According to the number of articles mentioned in blogs and news, PlumX and Altmetric.com show a similar proportion with regard to blogs (A=6%; P=5.7%), while Altmetric.com has a bit more articles mentioned in news media (A=3.7%; P=2.8%). CED shows a similar proportion with 1.2% of blogs and 0.7% of news.

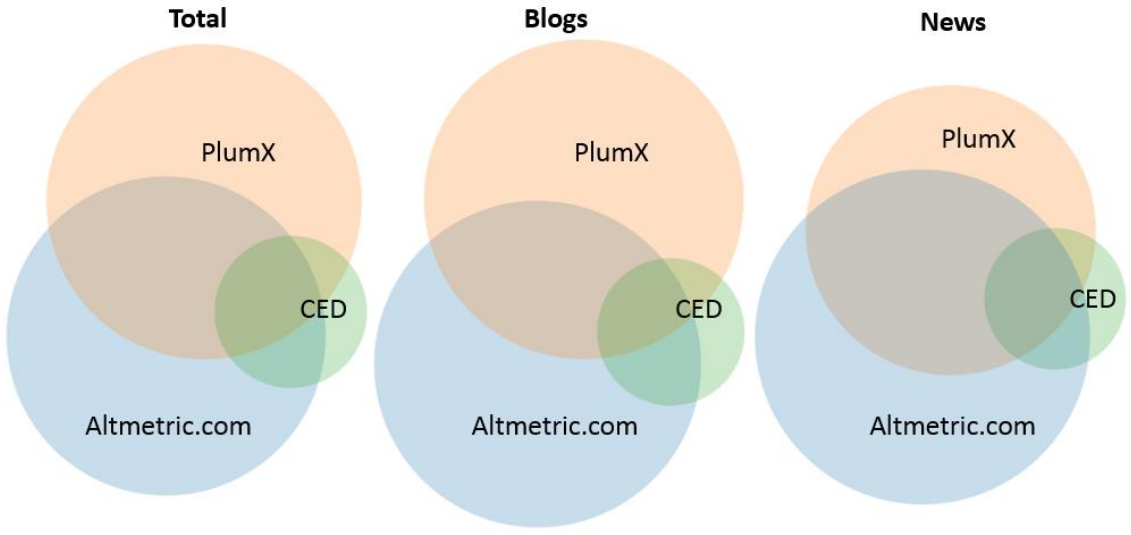

Figure 1. Venn diagram about the distribution and overlap of publications in the three major data providers.

Figure 1 shows the overlap of publications mentioned in the three altmetric providers according to news and blogs. The picture shows that there is low overlap between them (7.4%), being greater for news (7.9%) than for blogs (4.8%). The highest overlap is found between Altmetric.com and PlumX (28.8%), being more significant for news (36.8%) than for blogs (21.1%). These differences between blogs and news could be to the number of media is lower than the number of blogs, which could cause a higher overlap.

Events

Another way, although much more complex, to benchmark altmetric providers is to compare the number of events. URL of each event was used as identifier because different news and blog posts can have the same title. However, the URL matching could be problematic because both Altmetric.com and PlumX use third parties to provide links about blogs and news. In the case of Altmetric.com, 39.7% of the news outlets links come from Moreover.com (now Lexis-Nexis), while 5.2% of the blog posts links in PlumX are provided by ACI Scholarly Blog Index (now ProQuest). These services use link resolvers, and therefore the real URL is hidden. This problem gets worse when both services are now disappeared, and the links, in some cases, do not work properly. In the case of Altmetric.com, 8,794 (74.1%) links were resolved, while any link from ACI in PlumX could be resolved.

In addition, it was detected that some mentions do not actually come from news and blogs. Instead, they are citations from other research publications. This problem is not relevant in Altmetric.com because they just represent 0.8%. The real problem occurs in PlumX where 23.7% of the mentions are in fact bibliographic citations, mainly from Hindawi (17.9%) and OMICS Publishing Group (3.1%). These mistakes were detected more frequently in blogs (93.5%) than in news (6.5%). These citations were removed to do a fair comparison.

Another problem is that Altmetric.com does not provide full information about some events. In certain publications, Altmetric.com only shows the first four events while keeping the remaining ones hidden. The size of this gaps can be observed if the total number of events and the sum of counts are compared (Table 2). Notice that many events can mention more than one article, and then the number of real events has to be less than the cumulative count. This difference is significant in the number of news, where Altmetric.com only shows a 62% of the total counts. However, the percentage of blog posts (83.7%) could be considered acceptable. In PlumX, the percentage of blogs (27.9%) and news (90.2%) are caused by duplicated events and the omission of bibliographic citations aforementioned. There is not differences in CED because this provider does not show counts, but raw data.

Table 2. Distribution of events in the three major data providers

	Altmetric.com	*PlumX*	*CED*	*Sample*
Events	25,395	18,405	2,529	45,526
% Sample	55.8%	40.4%	5.5%	
Blogs	13,604 (83.7%)	3,216 (27.9%)	1,530	17,136
Count	16,242	11,537	1,530	
% Sample	29.9%	7.1%	3.4%	37.6%
% Provider	53.6%	17.5%	60.5%	
News	15,637 (61.7%)	15,288 (90.2%)	1,000	29,271
Count	25,321	16,949	1,000	
% Sample	34.3%	33.6%	2.2%	64.3%
% Provider	61.6%	83.1%	39.5%	

According to the distribution of events across the three providers, Table 2 shows the number and percentage of blog posts and news in Altmetric.com, PlumX and CED. In general, results show that Altmetric.com (55.8%) collects more events than PlumX (40.4%), while CED only achieves to capture a small amount of events (5.5%). However, each provider shows different proportions between blog posts and news. Altmetric.com is the source that includes more blog posts in the sample (29.9%), followed by PlumX (7.1%) and CED (3.4%). Whereas, Altmetric.com (34.3%) and PlumX (33.6%) captures a similar proportion of news, and much more than CED (2.2%). With regard to the proportion of blogs and news in each provider, the sample shows a higher proportion of news (64.3%) than blog posts (37.6%). PlumX is the service that presents a more disproportionate distribution (blogs=17.5%, news=83.1%), followed by CED (blogs=60.5%, news=39.5%). Altmetric.com is the service that shows a more equilibrate proportion (blogs=53.6%, news=61.6%).

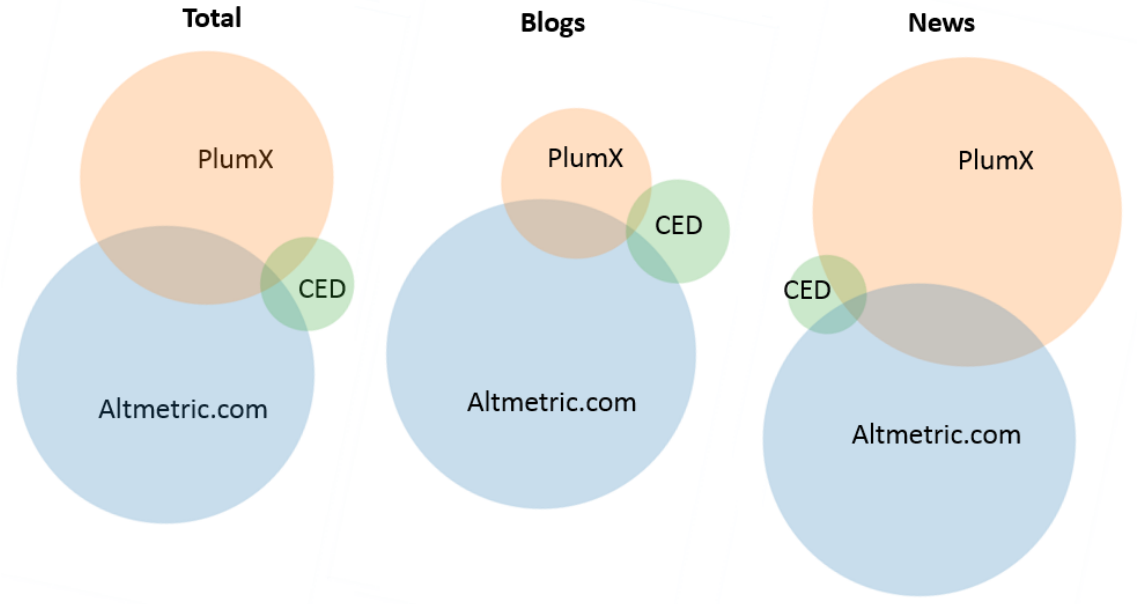

Figure 2. Venn diagram about the distribution and overlap of events in the three major data providers.

Figure 2 shows the overlap of blogs and news mentions between the three altmetric providers. Unlike Figure 1, the overlap of events is even lower than the overlap of publications. In general, only 0.5% of the events are simultaneously gathered by Altmetric.com, PlumX and CED, being the greatest overlap between Altmetric.com and PlumX (7.7%). According to blog posts, the overlap is diminished with only 0.3%, where Altmetric.com and PlumX gather together 5.5% of the blog posts. In the case of news, the overlap is slight higher (0.4%), being again Altmetric.com and PlumX the services that show the highest overlap (7.7%). This very low overlap among events evidences that the mention of articles in blogs and news is infrequent and they appear in a very varied range of sources that are not completely covered by all the altmetric providers together.

Sources

A third way to compare the altmetric providers is analysing the number of distinct sources that publish the blog posts and the news. Web domains and names of the sources were revised to merged duplicated sources. For example, blogs that change from web servers (botany.one, aobblog.com) or that have several domains (academiclifeinem.com, academiclifeinem.blogspot.com) were merged. Blogs hosted in the same platform such as *AGU Blogosphere* (blogs.agu.org) and *LSE Blogs* (blogs.lse.ac.uk) were distinguished. News media

with different languages editions (for example, *CNN* and *CNN en Español*) were differentiated as well, because it is assumed that the content is different in those editions.

Table 3. Distribution of sources in the three major data providers

	Altmetric.com	*PlumX*	*CED*	*Sample*
Sources	3,856	3,255	1,263	6,837
% Sample	56.4%	47.6%	18.5%	
Blogs	2,582	860	960	3,810
% Sample	37.8%	12.6%	14.0%	55.7%
% Provider	67.0%	26.4%	76.0%	
News	1,408	2,496	310	3,387
% Sample	20.6%	36.5%	4.5%	49.5%
% Provider	36.5%	76.7%	24.5%	

Table 3 details the number and percentage of sources found in the sample. Overall, 6,837 sources were detected. Altmetric.com contains 3,856 (56.4%) different sources, followed by PlumX with 3,255 (47.6%) and CED far away with 1,263 (18.5%). In general, the proportion of blogs 3,810 (55.7%) and news outlets 3,387 (49.5%) is rather equilibrated. However, this balanced distribution is broken when blogs and news are distinguished. According to blogs, Altmetric.com includes 2,582 (37.8%) sources, more than double that PlumX with 860 (12.6%) and CED with 960 (14%). This result emphasizes the good coverage of blogs by Altmetric.com. However, if the proportion of news media is observed, PlumX includes 2,496 (36.5%) sources, whereas Altmetric.com indexes 1,408 (20.6%) media and CED only 310 ones (4.5%). These differences could be due to the special coverage of the US local media by PlumX (Elsevier, 2018).

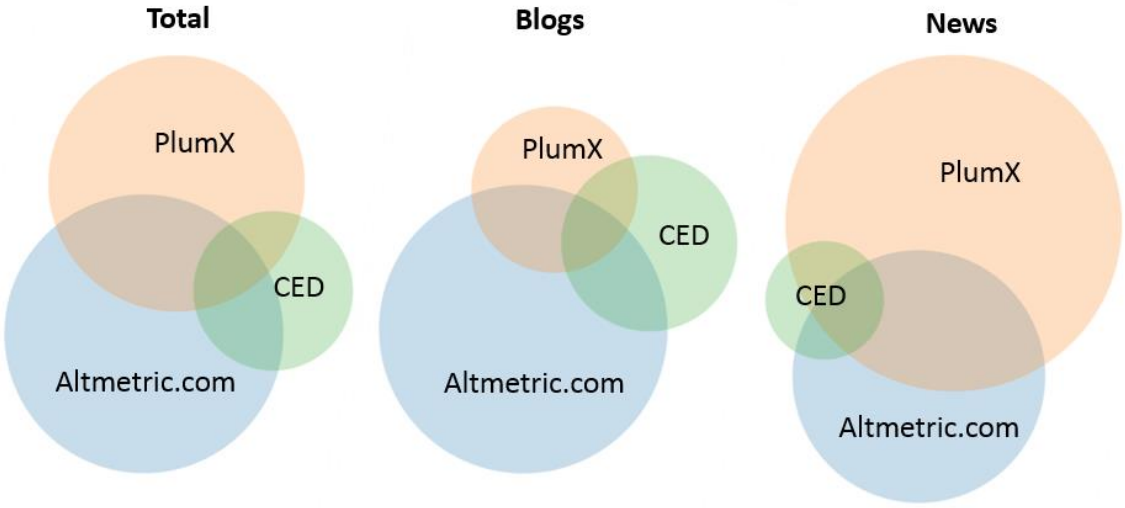

Figure 3. Venn diagram about the distribution and overlap of sources in the three major data providers.

Figure 3 shows the overlap of blogs and news sources between the three altmetric providers. Overall, the overlap between the three providers still is low (4.1%), being the couple Altmetric-PlumX which has the largest overlap (16%) and PlumX and CED are the aggregators that shares less sources (5.6%). According to blogs, the overlap between the three providers is only 1.6%, being again Altmetric.com-PlumX the pair that is more overlapped (5.7%) and PlumX and CED the services that share less sources (3.4%). News presents a little more overlap than blogs (1.8%), being larger between Altmetric.com and PlumX (9.7%). However, CED and Altmetric.com are the providers that have less news in common (2.3%).

Discussion

The results on the detailed coverage of Altmetric.com, PlumX and CED of blog posts and news have showed a great disparity between the indexation of articles, the capture of events and the coverage of sources. PlumX and CED index almost all the articles of the sample (99%), while Almetric.com only covers 34%. The cause of this low coverage is unknown and it could be due to publisher's agreements or that Altmetric.com only indexes articles that have some social event. Precisely, this last reason could be the most probable one, as Altmetric.com is the platform that has more articles mentioned in blogs and news media (7.4%), followed by PlumX (7.2%) and CED (1.7%). Results show that there is almost the double of articles cited in blogs (9.9%) than in news (4.8%), which suggests that the blog post format is more suitable for the discussion of new academic results (Shema et al., 2015; Ritson, 2016), while traditional news format is used more to disseminate outstanding advances (Bubela and Caulfield, 2004; Suleski and Ibaraki, 2010). In this sense, Altmetric.com is the service that more articles indexes in both type of events (blogs=6%, news=3.7%), followed by PlumX (blogs=5.7%, news=2.8%), and CED (blogs=1.2%, news=0.7%).

One of the most relevant facts in this study is that it is not based on counts but on events. This element had made possible to compare data providers according to their coverage of blogs posts and news. In this sense, it is interesting to emphasize that Altmetric.com is the service that captures more events (55.8%) and from more different sources (56.4%), with a well-adjusted distribution between blogs (29.9%) and news (34.3%). However, PlumX presents a more imbalanced proportion, with a very low percentage of blog posts (7.1%) and sources (12.6%) opposite to a good coverage of news (33.6%) and news outlets (36.5%). This last percentage is even above of Altmetric.com (20.6%). Two elements explain this biased distribution. First, PlumX has a deficient coverage of blogs because it includes research articles as blogs, which causes a miscount of blog posts. Second, PlumX shows a special coverage of local US media (TV and radio stations), which increases the mentions of articles from news (Elsevier, 2018). Finally, CED remains in a secondary role because it shows much lower figures than the two before services. For example, CED only covers 5.5% of the events, being 3.4% of blog posts and 2.2% of news. It is only remarkable that CED indexes more blog sources (14%) than PlumX (12.6%).

These important differences in the coverage of blogs and news is also reflected in the low overlap between services. Only 7.9% of the articles in the sample are indexed by the three data providers, being the highest overlap between PlumX and Altmetric.com (28.8%). This overlap is even more reduced when the events and sources are observed. Thus, 0.5% of the events and 4.1% of the sources are simultaneously gathered by the three platforms, being again the greatest overlap between Altmetric.com and PlumX (events=7.7%; sources=16%). This low overlap is consequence of the low number of scholarly results being commented in blogs and news (11.7%), which cause a great spreading of the mentions in a wide range of sources. In fact, the study gathers less than 4.000 sources for Altmetric.com and 3.200 for PlumX, a very low number if they are compared with the official figures (11,000 blogs and 1,300 news outlets for Altmetric.com and 10,000 blogs and 55,000 news sources for PlumX). However, all these disparities do not do anything but confirm that only one provider is not enough to observe the social impact of a publication and it is necessary the employment of several aggregators to undertake any reliable altmetric study.

However, this study goes beyond the event counts and, for first time, compares the content of the blog posts and news. This approach has disclosed import inconsistence in the count of these metrics. The distinction between blogs and news is not explained and many sources are classified in both groups. This happens in 3.5% of sources in Altmetric.com and 3.2% in PlumX. Although this percentage is low, it is indeed significant and it could influence the final count of blogs and news. However, without a doubt, the most important problem is caused by

the inclusion of bibliographic citations as blog mentions. This problem is insignificant in Altmetric.com (0.8%), but in PlumX is a serious issue because 23.7% of the blog mentions are, in fact, bibliographic citations. This result questions the reliability and accuracy of PlumX as data provider, concretely with regard to blogs.

The way in which the data were captured leads us to consider some important limitations when the results are interpreted. The most important limitation is that Altmetric.com, concretely Altmetric Explorer, does not show, in some cases, all the mentions that an article receives. This problem is especially significant in news where approximately 38% of the events were not retrieved. This problem can distort the results about coverage and overlapping, mainly, in the cases of events and sources. Another limitation is that many of the mentions come from disappeared services (Moreover.com and ACI Scholarly Blog Index), which caused that many of the events (5.2% of the blog posts in PlumX and 19.8% of the news in Altmetric.com) could not be verified and therefore compared with the other services. This technical problem would be able to influence the before figures about overlapping. However, this problem introduces the issue of the ability of these services to be audited by an independent organization that verified the counts that they publish (NISO, 2016). Anyways, future studies that test the coverage and overlapping between altmetric providers are welcome.

Conclusions

Several conclusions can be drawn from the results. First, PlumX and CED are the services that index more publications, reaching almost the totality of the sample. However, Altmetric.com is the service that gathers more documents mentioned in blogs and news (7.4%), closely followed by PlumX (7.2%) and CED (1.7%). The overlap between the altmetric providers is low (7.4%), being the greatest one between Altmetric.com and PlumX (28.8%).

Second, Altmetric.com is the platform that more blog and news events captures (55.8%), followed by PlumX (40.4%) and CED (5.5%). The proportion between blogs and news is balanced in Altmetric.com (blogs=29.9%; news=34.3%) and CED (blogs=3.4%; news=2.2%). However, PlumX has an important gap in the coverage of blog posts (7.1%) due to the miscounting of bibliographic citations as blog mentions. The overlap between platforms according to events is lower than according publications (0.5%), being the greatest for the couple Altmetric.com-PlumX (7.7%).

Third, according the sources of the events, Altmetric.com (56.4%) collects more distinct sources than PlumX (47.6%) and CED (18.5%). Again, Altmetric.com covers more blogs (37.8%), due to its exhaustive list of blogs. Whereas, PlumX highlights covering news outlets (36.5%), caused by the special coverage of local US media by its news provider, Newsflo. The overlap between the three services is again low (4.1%) and it makes evident the great amount of media that mention research outputs and the difficulty of gathering this information.

References

ACI (2016). ACI Scholarly Blogs Now Included in PlumX. Retrieved January 2, 2019 from: https://aci.info/2016/09/27/aci-scholarly-blogs-now-included-in-plumx/

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

Allen, P. (2017). World-Wide News Coverage in PlumX. Retrieved January 2, 2019 from: https://plumanalytics.com/world-wide-news-coverage-plumx/

Altmetric.com (2018a). How it works. Retrieved January 2, 2019 from: https://www.altmetric.com/about-our-data/how-it-works/

Altmetric.com (2018b). What outputs and sources does Altmetric track? Altmetric.com Support. Retrieved January 2, 2019 from: https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track-

Baessa, M., Lery, T., Grenz, D. & Vijayakumar, J. K. (2015). Connecting the pieces: Using ORCIDs to improve research impact and repositories. *F1000Research*, *4*, 195.
Bar-Ilan, J., Halevi, G. & Milojevic, S. (2018). Differences between Altmetric Data Sources–A Case Study. In *23rd International Conference on Science and Technology Indicators (STI 2018)* (pp. 699-702). Leiden: CWTS.
Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, *8*(4), 935-950.
Bubela, T. M. & Caulfield, T. A. (2004). Do the print media "hype" genetic research? A comparison of newspaper stories and peer-reviewed research papers. *Canadian Medical Association Journal*, *170*(9), 1399-1407.
Cave, R. (2013). Overview of the Altmetrics Landscape. In *Proceedings of the Charleston Library Conference*. Retrieved January 2, 2019 from: https://core.ac.uk/download/pdf/12982803.pdf
Chamberlain, S. (2013). Consuming article-level metrics: Observations and lessons from comparing aggregator provider. *Information Standards Quarterly*, 25(2), 4-13.
Elsevier. (2017). Elsevier Acquires Leading 'Altmetrics' Provider Plum Analytics. Retrieved January 2, 2019 from: https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-leading-altmetrics-provider-plum-analytics
Elsevier. (2018). Newsflo > Coverage and sources. Retrieved January 2, 2019 from: https://www.elsevier.com/solutions/newsflo/sources
Fraumann, G., Zahedi, Z. & Costas, C. R. (2015). What do we know about Altmetric.com sources? A study of the top 200 blogs and news sites mentioning scholarly output. In *The 2015 Altmetrics Workshop*. Retrieved January 2, 2019 from: https://altmetrics.org/wp-content/uploads/2015/09/altmetrics15_paper_19.pdf
Gumpenberger, C., Glänzel, W. & Gorraiz, J. (2016). The ecstasy and the agony of the altmetric score. *Scientometrics*, *108*(2), 977-982.
Jamali, H. R. & Alimohammadi, D. (2015). Blog Citations as Indicators of the Societal Impact of Research. *International Journal of Knowledge Content Development & Technology*, *5*(1), 15-32.
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*. Retrieved January 2, 2019 from: http://www.bibliometrie-pf.de/article/viewFile/205/258
Kraker, P., Lex, E., Gorraiz, J., Gumpenberger, C. & Peters, I. (2015). Research data explored II: The anatomy and reception of figshare. *arXiv preprint arXiv:1503.01298*. Retrieved January 2, 2019 from: https://arxiv.org/ftp/arxiv/papers/1503/1503.01298.pdf
Lindsay, J. M. (2016). PlumX from Plum Analytics: Not Just Altmetrics. *Journal of Electronic Resources in Medical Libraries*, *13*(1), 8-17.
MacLaughlin, A., Wihbey, J. & Smith, D. A. (2018). Predicting News Coverage of Scientific Articles. In *12th International AAAI Conference on Web and Social Media (*ICWSM 2018) (pp. 191-200). Palo Alto, CA: AAAI Press
Meschede, C. & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. *Scientometrics*, *115*(1), 283-297.
NISO (2016). Outputs of the NISO Alternative Assessment Metrics Project. NISO RP-25-2016. Retrieved January 2, 2019 from: https://groups.niso.org/apps/group_public/download.php/17091/NISO%20RP-25-2016%20Outputs%20of%20the%20NISO%20Alternative%20Assessment%20Project.pdf
Ortega, J. L. (2018a). The Life Cycle of Altmetric Impact: A Longitudinal Study of Six Metrics from PlumX, *Journal of Informetrics*, *12*(3), 579-589.
Ortega, J. L. (2018b). Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data. *Scientometrics*, *116*(3), 2123-2138.
Peters, I., Kraker, P., Lex, E., Gumpenberger, C. & Gorraiz, J. (2015). Research data explored: Citations versus altmetrics. In A.A. Salah, Y. Tonta, A.A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of the 15th International Conference on Scientometrics and Informetrics, Istanbul, Turkey* (pp. 172–183)
Peters, I., Kraker, P., Lex, E., Gumpenberger, C. & Gorraiz, J. (2016). Research data explored: an extended analysis of citations and altmetrics. *Scientometrics*, *107*(2), 723-744.

Plum Analytics (2018). Coverage: Expanding the World of Altmetrics. Retrieved January 2, 2019 from: https://plumanalytics.com/learn/about-metrics/coverage/

Ritson, S. (2016). 'Crackpots' and 'active researchers': The controversy over links between arXiv and the scientific blogosphere. *Social studies of science*, *46*(4), 607-628.

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

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

Suleski, J. & Ibaraki, M. (2010). Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media. *Public Understanding of Science*, *19*(1), 115-125.

Robinson-García, N., Torres-Salinas, D., Zahedi, Z. & Costas, R. (2014). Nuevos datos, nuevas posibilidades: Revelando el interior de Altmetric.com. *El Profesional de la Information*, *23*(4), 359-366.

Thelwall, M. (2018). Altmetric Prevalence in the Social Sciences, Arts and Humanities: Where are the Online Discussions?. *Journal of Altmetrics*, *1*(1). Retrieved January 2, 2019 from: https://www.journalofaltmetrics.org/articles/10.29024/joa.6/

Torres-Salinas, D., Gumpenberger, C. & Gorraiz, J. (2017). PlumX as a potential tool to assess the macroscopic multidimensional impact of books. *Frontiers in research metrics and analytics*, *2*, 5.

Torres-Salinas, D., Gorraiz, J. & Robinson-Garcia, N. (2018). The insoluble problems of books: What does Altmetric. com have to offer?. *Aslib Journal of Information Management*, *70*(6), 691-707.

Trueger, N. S., Thoma, B., Hsu, C. H., Sullivan, D., Peters, L. & Lin, M. (2015). The altmetric score: a new measure for article-level dissemination and impact. *Annals of emergency medicine*, *66*(5), 549-553.

Wass, J. (2018). Some thoughts on 'General discussion of data quality challenges in social media metrics'. *Joe's Blog*. Retrieved January 2, 2019 from: https://blog.afandian.com/2018/05/zahedi-costas-altmetrics/

Williams, C. (2015). Altmetric to provide expanded mainstream media tracking. *Altmetric blog*. Retrieved January 2, 2019 from: https://www.altmetric.com/blog/moreover/

Wong, E. Y. & Vital, S. M. (2017). PlumX: a tool to showcase academic profile and distinction. *Digital Library Perspectives*, *33*(4), 305-313.

Zahedi, Z., Fenner, M. & Costas, R. (2015). Consistency among altmetrics data provider/aggregators: What are the challenges?. In *The 2015 Altmetrics Workshop*. Retrieved January 2, 2019 from: https://altmetrics.org/altmetrics15/zahedi/

Zahedi, Z. & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PloS one*, *13*(5), e0197326.