Altmetric Counts from Different Sources – A Case Study of JASIST Articles Published Between 2001 and Mid 2017

Judit Bar-Ilan[1] and Gali Halevi[2]

[1]Department of Information Science, Bar-Ilan University, Israel
Judit.Bar-Ilan@biu.ac.il
[2]Icahn School of Medicine at Mount Sinai, New York, NY
halevi@mssm.edu

Extended abstract

The theme of this year's workshop is dependencies. It is, therefore seems appropriate to study the fundamental dependency that the bibliometric community has on data provided by third party vendors. The ability to develop accurate and scalable measures of the impact of scholarly output relies heavily on the various data types, both bibliometric and altmetric. Research has shown that the citations counts, for example, retrieved from Web of Science (WOS), Scopus and Google Scholar (GS) differ, thus resulting in different indicators assigned to both authors and publications. Differences in coverage practices and policies, for example, result in evaluative indicators to vary between databases (Bar-Ilan, 2008). There is a large body of literature that compares various databases and how their coverage differs while affecting their bibliometric indicators (Halevi, Moed & Bar-Ilan, 2017). However, there are not very many studies that examine whether such differences affect altmetric indicators. Therefore, this exploratory study aims to examine the issue of coverage differences and the manner by which they affect altmetric indicators. Our dataset consisted of all the articles and reviews (referred to as "articles") published in JASIST (Journal of the American Society for Information Science and Technology, 2001-2013 and Journal of the Association for Information Science and Technology, 2014-2017 issue7). Overall our database included 2,728 artifacts (2,666 articles and 62 reviews). We compared Mendeley counts from three data providers: Mendeley, Altmetric and PlumX. In addition, we retrieved tweets, blog posts, news media and Wikipedia mentions from Altmetric and PlumX – the major aggregators of altmetric signals. Data from all three sources were collected on the same day, on 29 June 2017.

Reader counts

Mendeley had the highest coverage, which was expected, because altmetric.com reports Mendeley reader counts only if the article receives at least one altmetric signal that is not from Mendeley. We were not able to find PlumX's data collection policy from Mendeley. It seems that they only collect data from Mendeley for publications that receive another altmetric signal (similar to Altmetric's policy), but there are numerous cases where there are other altmetric signals, but Mendeley reader counts are not reported even though they exist.

Mendeley reported reader counts for 2628 (96%) of the articles. This is quite remarkable, since this number is larger than the number of cited articles indexed by Scopus (2443, 90%). The reason we are comparing items with reader counts to cited items stems from the fact that a reader count means that a user was reviewing the artifact in one form or another. It could possibly be the author/s but since there were only 112 articles with less than four readers (the average number of authors is 2.4, median 2, maximum 20), it does

not seem like it is the case. Therefore, even if we remove these 112 articles, the coverage remains at 92%.

Altmetric reported Mendeley reader counts for 1,113 articles, while PlumX reported reader counts 1,721 articles. Therefore, there are considerable differences in the coverage that are likely to be the result of the data collection policies of the different aggregators. Figure 1 illustrates the overlap (or the lack of it) in coverage between the data providers. It is worth noting that there are 65 articles for which PlumX or Altmetric provide reader counts, while Mendeley itself does not have any such counts on its platform.

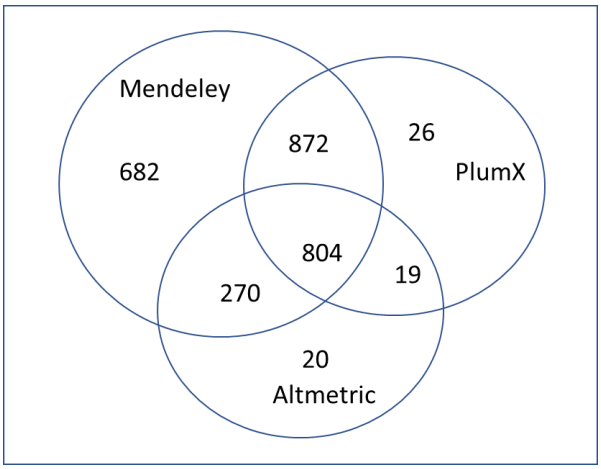

Figure 1: Number of articles for which Mendeley reader counts are provided

As can be seen from the above, there are not only differences in coverage, but there is some indication that there are differences in the number of readers reported (at least for the 65 articles with no readers on Mendeley, but with reader counts on Altmetric and/or PlumX). Table 1 displays the differences and the agreements between pairs of data sources. For each pair, we considered only cases where both sources reported reader counts.

X & Y	all	same	X more	Y more
X=Mendeley,Y=PlumX	1676	706 (42%)	522 (31%)	448 (27%)
X=Mendeley,Y=Altmetric	1074	875 (82%)	184 (17%)	15 (1%)
X=PlumX, Y=Altmetric	823	280 (34%)	309 (38%)	234 (28%)

Table 1: Agreements and disagreements in reader counts in absolute numbers and percentages

Differences in the reported reader counts may be explained by Mendeley's database update strategy, as explained by Gunn (2016). Mendeley reader count is immediately updated when a user adds to her personal library an item that existed on Mendeley

before, however adding items not indexed before and deleting items or entire libraries are only shown at the next periodic rebuild. Thus, Mendeley counts can decrease or increase after periodical rebuild of the database. The differences in the counts from the three sources are sometimes high as can be seen in Table 2.

X & Y	Max (X-Y)	Max (Y-X)	Average (X-Y)	Average (Y-X)
X=Mendeley,Y=PlumX	251	67	7.8	3.0
X=Mendeley,Y=Altmetric	225	41	17.8	6.3
X=PlumX, Y=Altmetric	95	165	4.2	5.7

Table 2. Maximum and average differences

Despite the differences demonstrated in the counts, Spearman correlations are very high and significant ($p<.001$): Mendeley-PlumX, $r=.954$; Mendeley-Altmetric, $r=.959$; and Altmetric-PlumX, $r=.970$. (Spearman correlations were computed because of the large standard deviations).

Twitter, Blogs, Wikipedia and Mainstream News – Almetric vs. PlumX

Social media indicators such as Twitter show considerable differences both in the coverage and the actual number of tweets reported for each article. The results are shown in Tables 3 and 4. For the 800 articles that have Twitter mentions on both platforms, the correlation is .864 ($p<.01$). This is lower that the correlation between Mendeley readers' counts.

	# articles with tweets	average # tweets per article	max # tweets per article
Altmetric all	980	7.2	402
PlumX all	824	9.2	723
Altmetric & PlumX - Altmetric counts	800	8.3	402
Altmetric & PlumX - PlumX counts	800	9.1	723
Altmetric only	180	2.5	60
PlumX only	24	4.8	48

Table 3. Number of articles with tweets, average and maximum tweet counts per article

	#A=#P	#A>#P	#P>#A	max A-P	max P-A
Altmetric & PlumX	336	225	239	69	707

Table 4. Differences in the number of reported tweets, when both Altmetric and Plum report tweet counts

Finally, Table 5 includes the number of articles mentioned in blogs, Wikipedia and mainstream news, and the number of mentions found on Altmetric and for PlumX. There are differences to note here as well. One explanation for the differences in counts for

blogs and mainstream media is that each aggregator covers different sources for such mentions. This is not the case for Wikipedia. Altmetric clearly states that it covers Wikipedia in English only. Although we could not find any clarification on PlumX, the fact is that its counts from Wikipedia are lower than that of Altmetric.

	# articles mentioned in blog	sum of blog mentions	# articles mentioned in Wikipedia	sum of Wikipedia mentions	# articles mentioned in mainstream media	sum of mentions in mainstream media
Altmetric	153	265	82	108	25	55
PlumX	37	62	77	103	7	9

Table 5. Differences in additional altmetrics

Conclusion

In this short paper, we demonstrated differences in altmetric counts between different data sources. The analysis shows that differences in counts of altmetric indicators exist in much the same way that bibliometric indicators differ between databases. Differences in coverage and frequency of updates influence such differences and should be noted when studying the impact of articles.

Acknowledgment

We thank Mike Thelwall for developing Webometric Analyst, which was used to collect data from Mendeley and Altmetric.

References

Bar-Ilan, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, *74*(2), 257-271.

Gunn, W. (2016). Comment #00632 - clarification/correction of Mendeley saves definition - NISO RP-25-201x-3, Altmetrics Data Quality Code of Conduct - draft for public comment.pdf. Retrieved from http://www.niso.org/apps/group_public/view_comment.php?comment_id=632

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. *Journal of Informetrics, 11*(3), 823-834.