Proposal of composed altmetric indicators based on prevalence and impact dimensions

José Luis Ortega

Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain, jortega@iesa.csic.es

Joint Research Unit Knowledge Transfer and Innovation, (UCO-CSIC), Córdoba, Spain

Abstract

The aim of this study is to introduce two groups of impact indicators, Weighted Altmetric Impact (WAI) and Inverse Altmetric Impact (IAI). WAI is based in weights from the contributions of each metric to different components or impact dimensions (Principal Component Analysis). IAI is calculated according to the inverse prevalence of each metric in different impact dimensions (TF/IDF). These indicators were tested against 29,500 articles, using metrics from Altmetric.com, PlumX and CED. Altmetric Attention Score (AAScore) was also obtained to compare the resulting scores. Several statistical analyses were applied to value the advantages and limitations of these indicators. Frequency distributions showed that each group of metrics (Scientific Impact, Media Impact and Usage Impact) follows power law trends although with particular patterns. Correlation matrices have depicted associations between metrics and indicators. Multidimensional scaling (MDS) has plotted these interactions to visualize distances between indicators and metrics in each dimension. The 2018 Altmetric Top 100 was used to distinguish differences between rankings from AAScore and the proposed indicators. The paper concludes that the theoretical assumptions of dimensions and prevalence are suitable criteria to design transparent and reproducible impact indicators.

Keywords: Altmetrics; Weighted Altmetric Impact; Inverse Altmetric Impact; Altmetric Attention Score; Bibliometrics

Introduction

In the scientific literature, sometimes, it is common to equally use the terms metric and indicator as synonymous. However, both terms are slightly different. A metric is a measure unit that quantifies a magnitude. It informs us how big or small is a systematic observation. An indicator, on the contrary, is a measure that indicates something or gives information about something (Cambridge Dictionary, 2019). This means that an indicator is an instrument to value the performance of a system, while the metric is just a quantifier of events.

In bibliometrics, this distinction is necessary to distinguish between easy counts of citations or papers and elaborated and normalized measurements that express the impact of publications (i.e. Field-Weight Citation Impact), authors (i.e. h-index) or research journals (i.e. Journal Impact Factor). These indicators attempt to normalize the measure by size elements (i.e. citations per

publication) or external variables (i.e. disciplines, publication date) that could distort the meaning of an observation.

Perhaps, this differentiation is even more important in Altmetric studies, where the concept of metric is employed to refer to the count of events occurred to articles on the Web. Thus, we talk about readers, blog mentions or tweets as alternative metrics. Whereas, the idea of indicator is often referred to compound measurements that aggregate the counts of one or several metrics, normalizing by any size factor (i.e. number of publications). In this sense, many of the initiatives to create altmetric indicators have dealt with commercial platforms (Altmetric.com, ResearchGate) that have designed simplistic all-in-one indicators based on arbitrary weights and the addition of unalike metrics (Gumpenberger et al., 2016; Orduña-Malea et al., 2017).

However, the absence of rigorous academic altmetric indicators based in theoretical assumptions could be motived by two important problems related to this type of metrics. The first one is the presence of multiple and different metrics whose meanings are in some cases very different or opposed. The mixture of patent citations with tweets and blog mentions only can produce that the meaning of the indicator is vague and imprecise. A prior analysis is needed on the specific meaning of each metric and how it is related to other measurements. This would make possible to design more robust indicators, taking into account similarities among metrics. The second problem is that many of these events have very different prevalence. Whereas more than 90% of the articles are read in Mendeley, a very reduced number of them are mentioned in news (4-5%) (Meschede and Siebenlist, 2018; Ortega, 2019b) or policy documents (0.5%) (Yu et al., 2019). The reason of these differences are that while the effort to retweet a document is minimal (and so much more frequent), to mention a paper in a blog requires a great compositional effort. These different behaviours and meanings should be taken into account when a composite indicator is designed.

This paper attempts to propose several altmetric indicators that fill this gap, considering the results of theoretical studies on the behaviour of these alternative metrics. Specially, considering the different nature and meaning of each metric and their likelihood of appearance.

Literature Review: Indicators

Current literature includes some examples of indicators that measure the impact of alternative metrics. The most common option has been to extrapolate bibliometric indicators to altmetrics as a way to fairly compare differences between both types of indicators. For example, Zahedi et al. (2017) used Mean Readership Score as equivalent to the Mean Citation Score to compare Mendeley readers with citations. Thelwall (2017a) and Li and Thelwall (2012) used the same criterion to compare the same metrics (Reader mean and Citation mean). In this same line, Bornmann and Haunschild (2015) employed a version of the h-index to build the T factor, an indicator to measure the impact of authors, journals or organization on Twitter. T factor is the same as the number of *n* articles with more or equal *n* number of tweets. These same authors (Bornmann and Haunschild, 2016) also proposed the Twitter Index. This indicator ranks research journals setting a threshold of 80% articles mentioned on Twitter. From this cut-off, percentiles are calculated to sort each journal. Its objective is to avoid differences between disciplines. ReadMeter, an extinct altmetric aggregator, adapted two bibliometric indicators for authors (the H-Index and the G-Index) and redefined these metrics by using bookmarks instead of citations (Cave, 2012). A more elaborate approximation was done by Thelwall (2017b), who designed the Mean Normalised Log-transformed Citation Score (MNLCS) to normalize the impact of scientific citations, Mendeley readers and Wikipedia citations.

However, these indicators are designed to work with only one variable or metric. Other studies have tried to measure the global impact of altmetrics using compound indicators. Alhoori and Furuta (2014) created the Journal Social Impact, an addition of social networks and media mentions to the articles of a journal divided by the number of items. Hassan and Gillani (2016) proposed the alt-index to measure the social impact of authors. Calculated as the h-index, this indicator aggregates a range of weighted altmetric mentions to publications, counting the number of n articles with more or equal n number of altmetric mentions. Hassan et al. (2017) used ASNIP, an altmetric derivation of the Source Normalized Impact per Paper (SNIP) in which the total altmetric counts are normalized by field. These approaches are also based in previous bibliometric indices (i.e. h-index, Journal Impact Factor) but without considering the differences of meaning and context between these metrics. For that reason, the simple translation of bibliometric indicators to altmetrics could be confusing and imprecise.

From the commercial world, the most well-known indicator is the Altmetric Attention Score (AAS) of Altmetric.com. This indicator is calculated adding all the metrics that are captured by Altmetric.com, excepting Mendeley readers, CiteULike saves (these sources do not allow to audit these metrics) and Dimensions citations (citations are not considered an alternative metric). These metrics are weighted according to arbitrary weights and several partial criteria are used to value different events and sources (Altmetric Support, 2019). This indicator presents important limitations because it combines different metrics in the same count as well as the weights and criteria are not based on scientific evidences (Gumpenberger et al., 2016; Mukherjee et al., 2018). Another issue is that the values are rounded off, causing important miscounts when the scores are low. This score has been tested by different authors, finding poor correlations with bibliometric indicators (Wang et al., 2014; Ruan et al., 2018; Huang et al., 2018).

ImpactStory is another altmetric provider that incorporates some indicators (*Achievements*) that measure the performance of author profiles in different social networks and web sites. Profiles are ranked regarding to different criteria and they get an achievement when they exceed a threshold. However, many of these ranking criteria are *sui generis* and are barely related to scientific impact, quality or prestige. For example, *Follower Frenzy* is when someone with more than 8,000 followers tweets an author's article or *Hot Steak* is when a paper has been commented for two months.

However, this review of altmetric indicators evidences that some of these approaches are simplistic because they make a literal translation of bibliometric indicators to altmetrics without understanding the intrinsic differences between altmetrics and bibliometrics. This cause a lack of scientific rigour when the indicators are designed, blending metrics with difference meaning and without considering the low prevalence of some of these metrics. On the part of commercial view, the main problem is that their indicators are not transparent and, therefore, not reproducible.

Objectives

The principal aim of this paper is to present and test two groups of altmetric indicators: Weighted Altmetric Impact (WAI) and Inverse Altmetric Impact (IAI). Both indicators attempt to solve two problems in altmetrics: (1) the use of several metrics in a compound indicator and (2) addressing differences in the prevalence and meaning of metrics. Research questions are:

- Are the theoretical assumptions of dimensions and prevalence suitable to design altmetric indicators?
- Is it possible that the proposed indicators could distinguish among different types of impacts?
- Can the proposed indicators show balanced information of the employed metrics?

Indicators

The main characteristic of the proposed compound indices is that they are based on weights that take into account meaning and prevalence. Thus, an indicator that aggregates several metrics needs to justify which metrics are used and what is the weight of each one in the final formula. Due to this, we need, first of all, to classify every metric in *dimensions* according to their significance and establish weights according to the *prevalence* or frequency of appearance.

Dimensions

Dimension is defined as the set of metrics that have a similar meaning and express the same impact in a precise environment. The meaning of similarity is defined as the proportion of shared information between two metrics, and environment is the place where these metrics have meaning and value. For example, a citation is valuable in a scientific environment while a tweet is relevant in a media environment. Lastly, impact is understood from a thermodynamic perspective and it alludes to the degree in which an object can change the state of a system. Thus, a publication impacts when it provokes different reactions (citations, mentions, saves, views, etc.) in a knowledge system (academia, media, social networks, etc.).

Recent literature has employed several multivariate statistical technics to classify metrics according to the information that their share. This classification enables the identification of impact dimensions. Zahedi et al. (2014) were the first ones that attempted to classify altmetric and bibliometric counts using factor analysis. Their results showed that tweets are uncorrelated with bibliometrics indicators, which demonstrated that Twitter mentions do not correspond to the same dimension of the bibliometric indicators. But the most interesting result is that Mendeley readers and Wikipedia citations are classified along citations, instead altmetrics. This result suggests that Mendeley readers and Wikipedia citations belong to the dimension of the scientific impact (Maflahi and Thelwall, 2016; Thelwall, 2017c, 2018). Later, these same authors used the same multivariate analysis to group several metrics in citation metrics, social network metrics and media metrics (Costas et al., 2015). These results were confirmed by Ortega (2018a), who defined three types of components: Media Impact (social networks and media), Usage Impact (views and downloads) and Scientific Impact (citations, bookmarks and reads). This last approach includes usage metrics because he takes data from different providers (Altmetric.com, PlumX and Crossref Event Data). Departing from this last classification, three different dimensions are defined relating to impact (Table 1).

Dimension	Metrics	Definition
Media Impact	Facebook	Number of mentions of a publication in Facebook's pages
	Blogs	Number of mentions of an publication on Blogs entries
	News	Number of mentions of a publication on News outlets
	Reddit	Number of mentions of a publication on the comments of Reddit
	Tweeters	Number of mentions of a publication on different tweets and retweets of Twitter

	Linkouts	Number of clicks on a link to download a full-text paper
Usage Impact	Abstract views	Number of accesses to the Abstract page of a publication
	HTML views	Number of accesses to the full-text document
Scientific Impact	CiteULike	Number of times that a publication is bookmarked on CiteULike
	Readers	Number of saves of a publication in the personal library of Mendeley users
	Citations	Number of mentions of a publication in the references of other research publications
	Wikipedia	Number of mentions of a publication in Wikipedia articles

Table 1. The twelve most frequent (prevalence) and contributory metrics grouped by impact dimensions

Table 1 only shows the most representative metrics due to their frequency and to the amount of information that they contribute to each dimension (Ortega, 2018a). Thus, Media Impact groups metrics from social networks and media, which inform about the impact of a research output in the public opinion; the Usage Impact gathers metrics related to the use that a person makes about an academic document; and finally, the Scientific Impact includes citations and reads and it is associated to the traditional bibliometric impact, but including Mendeley readers and CiteULike saves as new metrics linked to the scientific impact.

Prevalence

Another important factor to consider is the different prevalence of some metrics. Many studies have evidenced different frequency of appearance of many of these metrics. Thus, for example, whereas more than 90% of the scholarly outputs have at least one Mendeley reader (Didegah et al., 2018; Bar-Ilan et al., 2019; D'Angelo and Di Russo, 2019), the percentage of mentioned papers on blogs and news are very low (Robinson-Garcia et al., 2019; Ortega, 2019b). This very different frequency of appearance suggests that these metrics require different effort to be created, and therefore, they have distinct importance and reach. Holmberg (2015) claimed that the impact level of a metric is inversely proportional to the reach and prevalence of this one. Thus, a metric that has an elevated frequency has lower impact than a metric that occurs less times. For example, a tweet on a research paper has low impact because is very frequent and it requires less effort to be created. However, a mention in a blog or news outlet is less frequent because it needs more work, but it achieves higher impact. In a similar way, Qiu and Yu (2015) introduces a stratified pyramid-shaped scheme in which each metric is ranked by their frequency and cost. They establish a Perception level related to usage metrics (views, downloads) that forms the basic use of research papers; then the Social media level is related to actions in social media addressed to the dissemination; and the Application level is about the specific use of research results to produce new knowledge (blog posts, citations, news, etc.). Using a quantitative approach, we take the assumption that the impact or importance of a metric is inversely related to the frequency of appearance into a bibliographic corpus.

Weighted Altmetric Impact (WAI)

WAI is an indicator that attempts to summarize different types of altmetric impacts in only one value, taking into account dimensions and prevalence. Therefore, the metrics that correspond to each different altmetric impact should be previously defined. We call Weighted Media Impact (WMI) when we use only metrics related to the dimension Media Impact, and Weighted Usage

Impact (WUI) when usage metrics are considered. Formally, the WAI of a research output (i) is the addition of the values of each metric (a, b, c, etc.) multiplied by the weight of each metric (w_i).

$$WAI_i = a_i(w_a) + b_i(w_b) + c_i(w_c)$$

The weight of each metric is defined from a Principal Component Analysis (PCA). This multivariate statistical technique allows to know the contribution of the variables to each component. To calculate the contributions, this study employs a sample of 3,793 journal articles published in 2013 and retrieved in 2018 (Ortega, 2018a). The reason of using only articles from 2013 is that the impact of a document is time-dependent. This restriction avoids any bias caused by documents of different ages that accumulate different amount of mentions.

Metric	Number of events	Articles with events	Articles with events %	Events per article
Facebook	1,901	931	24.5%	2.04
Blogs	613	391	10.3%	1.57
News	1,248	319	8.4%	3.91
Reddit	41	30	0.8%	1.37
Tweeters	20,210	3,383	89.2%	5.97
Linkouts	119,156	2,778	73.2%	42.89
Abstract views	1,504,476	3,658	96.4%	411.28
HTML views	479,836	1,239	32.7%	387.28
CiteULike	626	347	9.1%	1.80
Readers	120,387	3,720	98.1%	32.36
Citations	72,349	3,507	92.5%	20.63
Wikipedia	302	180	4.7%	1.68
Total	2,321,145	3,793	100%	

Table 2. Descriptive statistics by metric of the sample used to calculate the PCA components.

Table 2 shows the number of events, articles with event and events per article by metric with the aim of show descriptive statistics. The metrics of the sample were taken from Altmetric.com (Facebook, Blogs, News, Reddit, Tweeters, CiteULike), PlumX (Linkouts, Abstract views, HTML views, Readers, Scopus citations) and Crossref Event Data (CED) (Wikipedia citations).

	Contribution		
	Media Impact	Usage Impact	Scientific Impact
Facebook	**20.137**	0.536	0.001
Blogs	**20.819**	0.016	2.554
News	**22.930**	0.064	1.238
Reddit	**13.469**	0.064	0.476
Tweeters	**16.203**	3.277	1.951
Linkouts	0.395	**33.013**	0.622
Abstract views	0.088	**41.435**	0.731
HTML views	0.049	**17.057**	0.023
CiteULike saves	2.495	0.375	**14.617**

Mendeley readers	0.242	4.137	**35.662**
Scopus citations	0.302	0.015	**38.390**
Wikipedia citations	2.871	0.011	**3.737**

Table 3. Contributions (%) of each metric to each component (PCA)

Table 3 shows the contributions of the metrics selected in Table 1 to each component. Other metrics such as citations from policy documents, peer review sites or syllabi were excluded from the model due to the low number of observations and zero variance, which it makes impossible to calculate correlations between those variables. Because every metric contributes to some extent to each component or dimension, only significant contributions (in bold more than 5%) were selected and the weight of each metric was recalculated to sum to 1. It is interesting to note the scores of Wikipedia citations, which contribution to Media Impact (2.87%) is closed to Scientific Impact (3.74%). These similar and low contributions suggest that Wikipedia citations have an imprecise meaning between Media and Scientific Impact.

	Weights		
	Media Impact	Usage Impact	Scientific Impact
Facebook	**0.22**		
Blogs	**0.22**		
News	**0.25**		
Reddit	**0.14**		
Tweeters	**0.17**		
Linkouts		**0.36**	
Abstract views		**0.45**	
HTML views		**0.19**	
CiteULike saves			**0.16**
Mendeley readers			**0.39**
Scopus citations			**0.42**
Wikipedia citations			**0.04**

Table 4. Assigned weights to each metric according to their contribution to each component

Table 4 presents the contributions after transformation. In this case, the summation of the weights in each component or dimension is 1. This introduces the limitation that the weight of each metric depends on the number of metrics in the component and their contributions. For example, if a new metric is considered in the Scientific Impact dimension (e.g. policy documents), all the weights have to be recalculated to distribute the contribution among the new metrics. According to our model, the WUI of an article with 10 tweets and one blog post will be WUI= 10*0.17 + 1*0.22=1.92, while a paper with 8 readers and 2 citations will have a WSI=3.96.

Inverse Altmetric Impact (IAI)

Inverse Altmetric Impact is a similar indicator to the Weighted Altmetric Impact because both come from the summation of weights and they are expressed according to dimensions. Thus, we talk about Inverse Usage Impact (IUI) when we use usage metrics, and Inverse Media Impact (IMI) if metrics from media are calculated. However, the weights of these indicators are based

on different criteria, the inverse contribution of each metric to its dimension. This concept comes from the assumption that a metric that is less frequent is more important than a metric very frequent. This is because a less frequent metric requires more effort and cost to be done (i.e. blog post, bibliographic citation) than other very frequent which, in many cases, is the results of an action or click (i.e. tweet, like) (Holmberg, 2015; Qiu and Yu, 2015).

This concept of inverse contribution recalls the principle of least effort (Zipf, 1949) in which the terms less frequent are the words with more significance, while the most frequent ones are articles and conjunctions that lack of meaning by themselves. This idea was used by Spärk Jones (1972) to formulate the Term Frequency / Inverse Document Frequency (TF/IDF), a statistical definition of the importance of terms in a corpus according to the inverse frequency of occurrence. Taking the TF/IDF formula into account, the inverse weight of a metric is

$$ w_i^m = A_i^m log\left(\frac{N}{D^m}\right) $$

where the weight (w) of a metric (m) in the publication (i) is the result of the value (A) of that metric (m) in the publication (i) multiplied by the logarithm of the total number of publications in the database (N) divided by the number of publications (D) that has the metric m. In other words, the frequency of an altmetric event (A^m) corresponds to term frequency (TF) and the logarithm of the inverse frequency of mentioned documents in the database ($log(N/D^m)$) concurs with inverse document frequency (IDF).

The main limitation of this formula is that N and D are not constant because they change when a new document is included in the system. Another problem is that these parameters are not always known. This would make it impossible to replicate the calculation of this indicator. A solution is to convert those values in a predefined constant from previous observations about the prevalence of each metric in different samples.

Metric	N/D_m Mean	Standard Deviation (SD)	Lower bound (95%)	Upper bound (95%)
Facebook	**4.68**	0.98	3.46	5.29
Blogs	**9.48**	3.52	5.77	11.38
News	**10.24**	3.70	6.26	13.89
Reddit	**56.59**	23.47	33.99	82.19
Tweeters	**1.21**	0.16	1.09	1.33
Linkouts	**1.79**	0.20	1.69	1.8
Abstract views	**1.18**	0.08	1.13	1.23
HTML views	**3.21**	1.35	1.79	4.25
CiteULike saves	**7.59**	2.53	5.4	10.19
Mendeley readers	**1.07**	0.05	1.03	1.12
Scopus citations	**1.29**	0.12	1.22	1.37
Wikipedia citations	**19.02**	13.87	5.71	29.89

Table 5. Mean, standard deviation and confidence intervals of N/D_m values from several previous studies.

Table 5 shows the average, standard deviation and confidence intervals of N/D_m values from the last published studies on the coverage of altmetric data providers (Fang & Costas, 2018; Meschede and Siebenlist, 2018; Ortega, 2018b; 2018b; Zahedi and Costas, 2018; Gorraiz et al.,

2019; Ortega 2019a; 2019b). Excepting Reddit (SD=23.47) and Wikipedia citations (SD=13.87), the less frequent metrics, all the values show suitable deviations. Taking these parameters into account, the weight of each metric can be calculated. Thus, the weight of Tweeters in a document with 10 tweets is $w^{Tweeters}$ = 10*(Log(1.21)) = 0.83 and the weight of Blogs with one blog mention is w^{Blogs} = 1*(Log(9.48)) = 0.97. The final result is the addition of the weights of each metric in the dimension, being the Inverse Media Impact (IMI)=$w^{Tweeters}$ + w^{Blogs} =1.8.

$$IAI_i = (w_i^a) + (w_i^b) + (w_i^c)$$

Comparative Analysis

The proposed indicators were tested in 29,500 articles from Altmetric.com, PlumX and CED from a previous study (Ortega, 2019a). This sample is independent form the sample of 3,793 articles used to compute the PCA (Table 3 and 4). The reason of selecting this sample is that its size is larger and their observations are independent from the data used to construct the indicators. Altmetric Attention Score (AAScore) was also include to compare its performance with the proposed indicators. This sample allows us to observe differences and similarities between indicators.
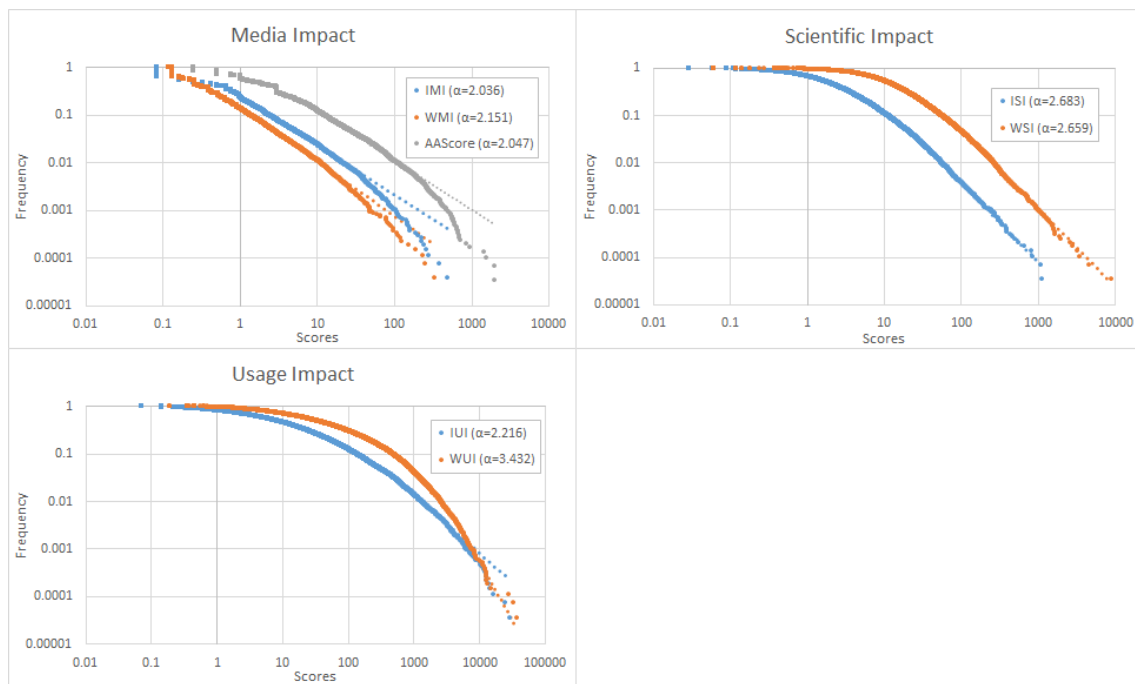
Distributions

Figure 1. Distribution and fit of the proposed indicators grouped by dimension (log-log)

Figure 1 shows the frequency distributions of the proposed indicators and AAScore in a log-log plot. The first impression is that the statistical behaviour of each indicator is strongly associated to its dimension. Thus, indicators based on Media metrics describe a very skewed trend with exponents between 2 and 2.1. Contrarily, indicators from usage metrics present an initial paused trend common in usage phenomena (Nielsen, 1997), in which the principle of least effort determines a small number of initial elections (Zipf, 1949). In other words, the usage of documents (visits to web pages, reads of articles, etc.) is rather random in the first stages of publication. Scientific indicators follow an intermediate pattern with a paused decay of the number of events by paper, followed by a sudden descent in the final part. Exponents are around

2.6. These different behaviours confirm that it is necessary to establish separate dimensions and differentiate the type of impact. According to the two types of calculation, weighted indicators present higher exponents than inverse indicators, which demonstrates that values from weighted indicators are more skewed that the inverse ones. AAScore, in comparison with the Media Impact indicators, follows a parallel trend with a similar exponent. This suggests that indicators from the same dimension have a similar statistical behaviour.

Correlations
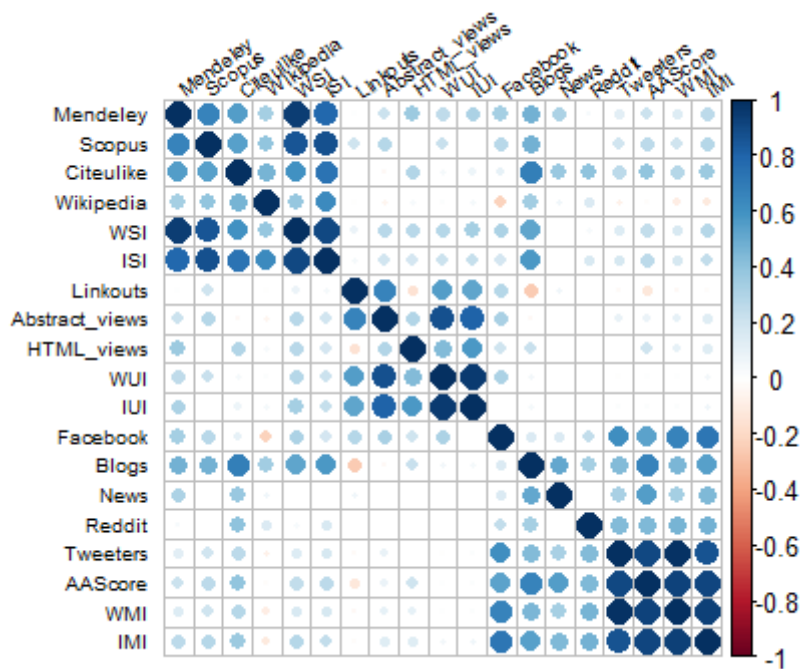
Figure 2. Spearman's rank correlation between metrics and indicators

Figure 2 presents the correlation matrix (Spearman's rank) of metrics and indicators. It shows how the correlations between metrics and indicators are grouped in three groups, Scientific Impact, Usage Impact and Media Impact. It is interesting to notice interactions between the three groups, specially the case of Blogs. This metric has high correlations with Scientific Impact metrics (CiteULike, $\rho=0.75$, p-value<.0001, and Mendeley readers, $\rho=0.53$, p-value<.0001) which suggests that the mention of research articles on blogs could be associated to the saving of publications on reference managers. Minor interactions were observed between usage metrics and Mendeley readers (Abstract views, $\rho=0.36$, p-value<.0001 and HTML views, $\rho=0.45$, p-value<.0001), evidencing the relationship between usage and saving. It is also interesting to see significant correlations between Scopus citations and Facebook ($\rho=.50$, p-value=.01) and Blogs ($\rho=.47$, p-value<.0001) that put in relationship the connection of the media metrics in the later citation impact (Thelwall et al., 2013; Shema et al., 2014).
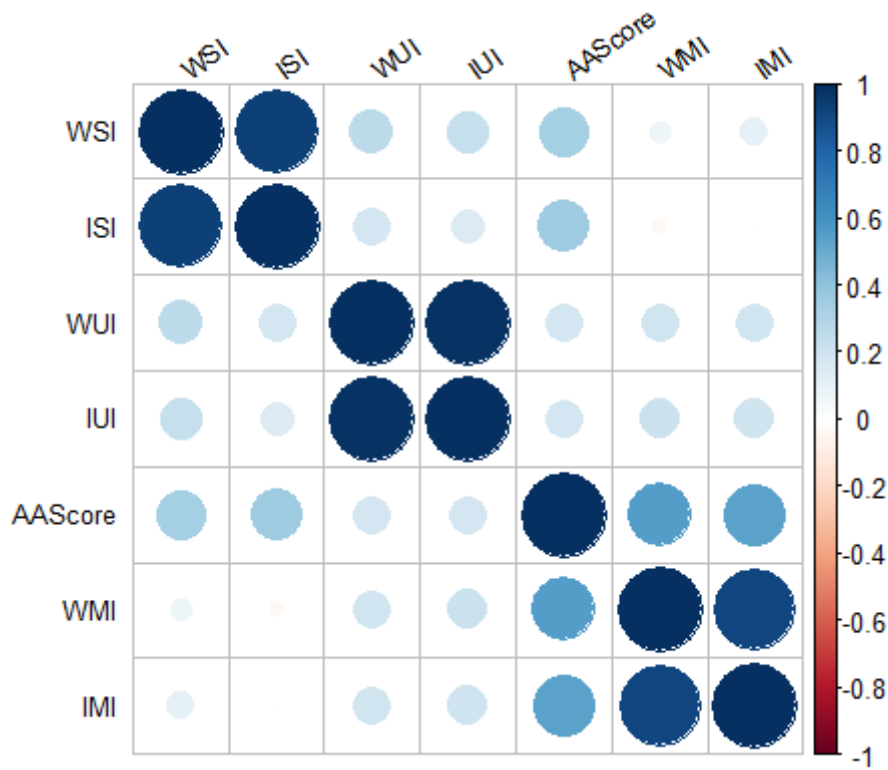
Figure 3. Spearman's rank correlation between indicators (p-value<.0001)

Figure 3 only shows correlations between the proposed indicators and the AAScore. The matrix presents more clearly how the three dimensions are differentiated and how the interactions between groups are low (ρ<0.30). The only exception is AAScore, which has a correlation of ρ=0.36 with ISI and ρ=0.34 with WSI, suggesting that the Altmetric.com's indicator captures a small fraction of the scientific impact. However, AAScore depicts a stronger relationship with media indicators (WSI, ρ=0.56 and ISI, ρ=0.54) due to these indicators share similar metrics and are oriented to the media impact.
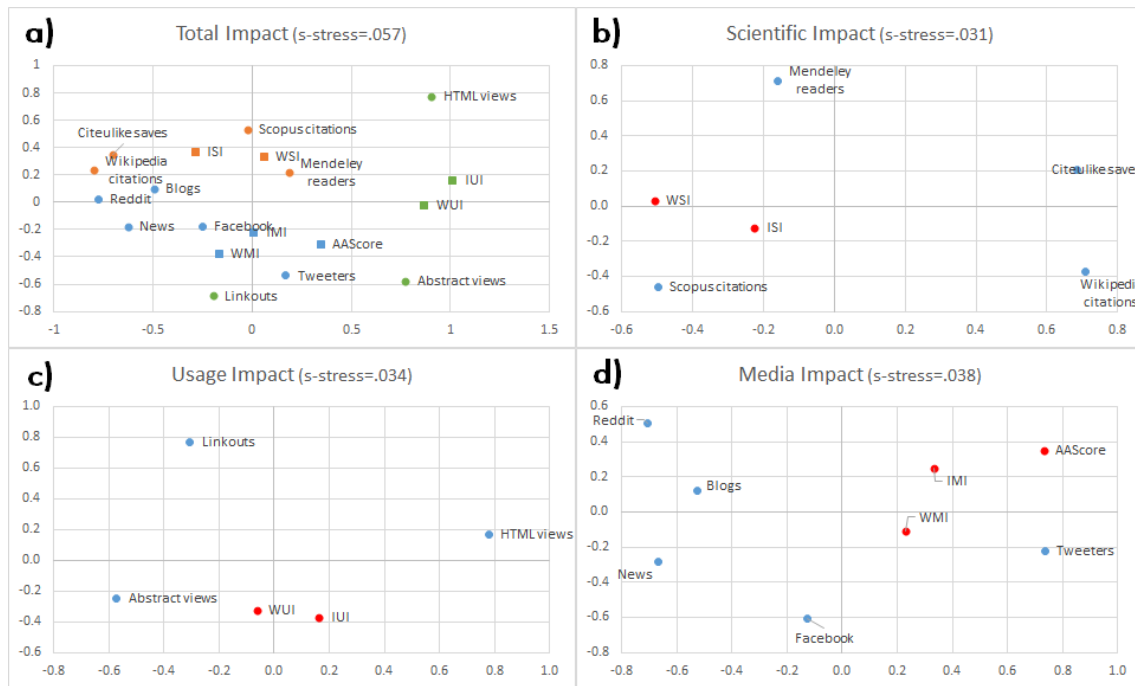
Figure 4. Multidimensional scaling (MDS) of the metrics in a log scale. Euclidean distance.

Figure 4 depicts the distance between metrics and indicators using multidimensional scaling (MDS). This multivariate technique enables to visualize a distance matrix based on correlations or similarities in a two or three-dimensional map. This mapping technique permits to observe in detail the proximities and similarities between indicators and metrics. Figure 4a displays all the metrics together to observe distances between groups. The graph shows that the groups are well-defined, which confirms that the metrics are associated to different types of impacts (Scientific Impact in orange, Usage Impact in green and Media Impact in blue). Figure 4a also shows how Blogs (Media Impact) and Wikipedia citations (Scientific Impact) move their groups closer, evidencing the ambiguity of Wikipedia and the connection of Blogs with reference managers (Fig. 2). Then again, the Usage Impact shows a great dispersion caused by the strong interactions of these metrics with Scientific and Media Impact dimensions.

Figures 4b, 4c and 4d shows the distances between metrics (blue) and indicators (red) into their respective dimensions. Both in Scientific Impact and Usage Impact, the proposed indicators occupy an intermediate position between the metrics, which evidences that these indicators contain a well-balanced information of every metric. This is more significant in the Media Impact, where the proposed indicators are located between Tweeters, the most frequent metric, and the remaining media metrics, being the proposed indicators equilibrate and representative indicators of the Media Impact. It is interesting to observe the position of the AAScore, very close to Tweeters. This confirms that this indicator is mainly based on the tweets, underrepresenting the information from other relevant metrics such blogs, news or Facebook mentions.

2018 top articles

A last analysis is to study the top 10 articles according AAScore and the proposed indicators as a way to point out differences between rankings at top positions.

Rank	Title	Journal	AAScore	WMI	IMI	WSI	ISI	WUI	IUI
1	Mortality in Puerto Rico after Hurricane Maria	New England Journal of Medicine	10441	1763.8	1450.1	159.6	22.8	357.5	816.3
2	The spread of true and false news online	Science	9816	1160.1	1199.2	688.9	93.1	1139.5	917.2
3	Alcohol use and burden for 195 countries and territories, 1990-2016…	The Lancet	7934	1182.8	1135.6	483.7	60.1	950.9	699.3
4	Trajectories of the Earth System in the Anthropocene	PNAS	6930	783.6	942.8	694.3	87.1	374.5	528.7
5	Weaponized Health Communication…	American Journal of Public Health	6438	1588.5	1299.6	117.5	17.8	3525.1	1261.8
6	Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015…	The Lancet Psychiatry	5988	1074.0	977.2	173.2	18.3	543.6	533.6
7	Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic	Scientific Reports	5594	763.5	834.6	383.1	42.1	462.9	520.2
8	Dietary carbohydrate intake and mortality…	The Lancet Public Health	5481	1109.4	972.6	198.2	23.8	2301.7	464.3
9	Complementary Medicine, Refusal of Conventional Cancer Therapy, and Survival Among Patients With Curable Cancers	JAMA Oncology	4835	714.6	625.6	79.1	9.2	117.1	372.1
10	The biomass distribution on Earth	PNAS	4822	779.4	683.9	589.0	77.6	39.6	356.7

Table 6. Ranking of the 2018 Altmetric Top 10 and according to the proposed indicators

Table 6 presents the first 10 articles by AAScore and their corresponding values according to the proposed indicators. The aim of this analysis is to observe the behaviour of those indicators in articles with an elevated number of events. Cells are coloured according to the ranking of each article in each indicator with the end of emphasizing ranking differences. Red corresponds to highest positons, while blue to lowest ones. WMI and IMI, indicators related to Media Impact, show similar positions as AAScore with minor differences. For example, "Trajectories of the Earth System in the Anthropocene", 4th according to AAScore, descends to the 8th position in WMI due to it is the second one with less tweets. That is the same with "Weaponized Health Communication…" that increases its position to the second place due to being the second most tweeted. These results show that IMI and WMI are more sensible to elevated number of tweets. WSI and ISI, Scientific Impact indicators, describe more significant differences with AAScore due mainly to the weight of Mendeley readers and Scopus citations. For instance, "Trajectories of the Earth System in the Anthropocene" is located in the first position of WSI because it is the most read article. The same happens with "The biomass distribution on Earth", the last one in AAScore but the 3rd one in WSI because it is the second one with more Mendeley readers. ISI, however, shows a higher weight of bibliographic citations, being "The spread of true and false news online" the first one according ISI because it is the most cited paper. This same happen with WUI and IUI, "Weaponized Health Communication…" is the first article according to these indicators because they are the publications with most Linkouts and Abstract views.

Discussion

This study introduces and analyses several indicators with the aim of improving the assessment of research articles, filling the existing gap on the development of specific altmetric indicators. The importance of these indicators stems from they are based on theoretical assumptions from the altmetric research, and not from arbitrary and non-transparent criteria. The first assumption is that not every metric provided by altmetric aggregators should be considered altmetrics or social media metrics. Other than citations, these sites provide a mix of metrics that have been

grouped as altmetrics, depicting a vague and imprecise view on the research impact. Research on this subject has demonstrated that these metrics are not homogenous and we cannot talk on only one type of impact. Based on previous analyses (Zahedi et al., 2014; Costas et al., 2015; Ortega, 2018a), this study proposes to initially distinguish between three types of impacts or dimensions. Scientific Impact expresses the impact of a research output in the own scholarly community through academic indicators such as citations and saved references; Usage Impact measures the use of scientific publications with reads, visits, downloads, etc.; and the Media Impact would be the core of altmetric, bringing together the social impact from media (blogs and news) and social networks (Twitter, Facebook, Reddit). The comparative study has proved that this distinction in dimensions is necessary due to several reasons: different statistical distributions, poor correlations between distinct dimensions (ρ<0.30) and significant distances between metrics from different types of impacts. These results allow to improve the altmetric research specifying the meaning of these metrics and identifying different types of impacts.

The second assumption is that the frequency of appearance or prevalence of these metrics are very different, and these differences are related to the reach (Holmberg, 2015) and the cost of producing those metrics (Qiu and Yu, 2015). Thus, less prevalent metrics are measures that require a great effort to be generated and have a high reach, whereas very frequent metrics are low cost counts that have short reach and scarce significance. These scope based on the prevalence has enabled to design two types of lineal or composed indicators, Weighted Altmetric Impact (WAI) and Inverse Altmetric Impact (IAI). Both indicators use different weights for each metric according to prevalence criteria. In the case of WAI, contributions to different PCA components are used, while in the case of IAI, a version of the TF/IDF formula is employed. The results have showed that this approach permits to build well-balanced indicators that include proportional information of every metric. This is perceived in the correlation matrix, where the indicators have high correlations with their metrics (ρ>0.32). MDS maps also display an intermediate position of these indicators regarding to their metrics.

But, perhaps, the most relevant advantage of this proposal is that these indicators are transparent, reproducible and based on theoretical criteria founded on scientific results (Hicks et al., 2015). The study includes the resulting weights and how have been calculated, which favours the utilization and reproduction of these indicators in any environment. The fact that they are supported by theoretical assumptions would reinforce the foundations of altmetrics and bibliometrics and encourage the public discussion and improvement of these indicators. In short, this study looks more for the theoretical discussion about how design suitable indicators from scientific foundations, than the presentation of definitive indicators for research evaluation.

Limitations

However, the implementation of these indicators presents some limitations. The first one is that they require previous information about the weights of each metric. In the case of WAI, contributions to each component are required. In IAI, it is before necessary to calculate N/D_m values. These values could not be constant and would change in subsequent analyses. Due to this, these weights should be a standard and their modifications must be validated by scientific consensus.

A second limitation is the inclusion of metrics in each indicator. In our examples, we have selected only the most significant metrics, motivated by the frequency and to the amount of information that they contribute to each dimension. However, these criteria could be modified

and new and relevant metrics such as policy documents and patent citations could be included to produce new dimensions or reinforce the existing ones. The fact that these indicators are linearly composed, favouring that new metrics can be included or removed without redefining the structure of the indicators.

A third limitation is derived from the unequal coverage of altmetric providers. Many studies have pointed important differences in the coverage of metrics and publications by these aggregators (Meschede and Siebenlist, 2018; Zahedi and Costas, 2018; Ortega, 2018b) which could bring different prevalence percentages regarding the used source for the calculations. In this case, we have used different sources to select the highest value of each metric but this does not ensure that these are the correct ones. In addition, coverage percentages could change in the time and some of these weights could be altered in next observations. For example, the recent defunct of Google+ (2019) has committed us to remove this metric for the study.

Another limitation is related to the theoretical presuppositions that drive these indicators. The concept of prevalence and its inverse relationship with the reach and cost of the mentions could be influenced by the usage and population of social networks. For instead, the low prevalence of Reddit posts could be more motivated because Reddit is a specialized network with few users than because their comments require more effort to be released. In the same manner, CiteULike bookmarks and Mendeley readers, two similar metrics, show a very different prevalence caused by the different number of users in each platform. These examples highlight the importance of controlling this variable when altmetric indicators are designed. It would be interesting and necessary to explore the way in which the number of users in these platforms can be used to counteract distortions in the prevalence.

Conclusions

Several conclusions can be drawn from the results. The theoretical assumptions of dimensions and prevalence have demonstrated that it is possible to design indicators supported by scientific results, giving more scholarly rigor, transparency and reproducibility to these measurements. The employment of statistical techniques such as PCA and TF/IDF has granted to these indicators more consistency and reliability.

This study has showed that there are different types of impacts and that it is necessary distinguishing between them to precise the meaning of the indicators. The definition of Scientific, Usage and Media Impact makes possible to group similar metrics according to their significance and permit to observe the influence of research outputs on new knowledge spheres.

The proposed metrics have achieved the aim of integrating the information of similar metrics in well-balanced measures. MDS maps and correlation matrices have shown that both WAI and IAI are well correlated with their metrics and they are located in an intermediate place between the individual metrics.

References

Alhoori, H., & Furuta, R. (2014). Do altmetrics follow the crowd or does the crowd follow altmetrics?. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 375-378). IEEE Press.

Altmetric Support (2019). How is the Altmetric Attention Score calculated? https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-

Bar-Ilan, J., Halevi, G., Milojevic, S. (2019). Differences between Altmetric Data Sources–A Case Study. *Journal of Altmetrics*, 2(1). https://doi.org/10.29024/joa.4

Bornmann, L., & Haunschild, R. (2015). t factor: A metric for measuring impact on Twitter. *arXiv preprint arXiv:1508.02179*.

Bornmann, L., & Haunschild, R. (2016). How to normalize Twitter counts? A first attempt based on journals in the Twitter Index. *Scientometrics*, *107*(3), 1405-1422.

Cambridge Dictionary (2019). Cambridge Dictionary. English Dictionary. https://dictionary.cambridge.org/dictionary/english/

Cave, R. (2012). Overview of the Altmetrics Landscape. Proceedings of the Charleston Library Conference. http://dx.doi.org/10.5703/1288284315124

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

D'Angelo, C. A., Di Russo, S. (2019). Testing for universality of Mendeley readership distributions. *Journal of Informetrics*, 13(2): 726-737. https://doi.org/10.1016/j.joi.2019.03.011

Didegah, F., Bowman, T. D., Holmberg, K. (2016). Increasing our understanding of altmetrics: identifying factors that are driving both citation and altmetric counts. In: *IConference 2016 Proceedings*.

Fang, Z., & Costas, R. (2018). Studying the Velocity Index for various Altmetric. com data sources. In 23rd International Conference on Science and Technology Indicators (STI 2018), September 12-14, 2018, Leiden, The Netherlands.

Gorraiz, J., Ulrych, U., Blahous, B., & Wieland, M. (2019). Monitoring the broader impact of the journal publication output on institutional level. Qualitative and Quantitative Methods in Libraries, 7(2), 293-308.

Gumpenberger, C., Glänzel, W., & Gorraiz, J. (2016). The ecstasy and the agony of the altmetric score. *Scientometrics*, *108*(2), 977-982.

Hassan, S. U., & Gillani, U. A. (2016). Altmetrics of" altmetrics" using Google Scholar, Twitter, Mendeley, Facebook, Google-plus, CiteULike, Blogs and Wiki. *arXiv preprint arXiv:1603.07992*.

Hassan, S. U., Imran, M., Gillani, U., Aljohani, N. R., Bowman, T. D., & Didegah, F. (2017). Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. *Scientometrics*, *113*(2), 1037-1057.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429.

Holmberg, K. (2015). Altmetrics for Information Professionals: Past, Present and Future. Oxford: Chandos Publishing

Huang, W., Wang, P., & Wu, Q. (2018). A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals. *PloS one*, 13(4), e0194962.

Qiu, J., & Yu, H. (2015). Stratifying Altmetrics Indicators Based On Impact Generation Model. *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*.
https://pdfs.semanticscholar.org/8fd1/fc122678c8afe725cbeba67319239f12d93c.pdf

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th international conference on science and technology indicators* (Vol. 2, pp. 451-551). Montréal, Canada: Science-Metrix and OST.

Maflahi, N., & Thelwall, M. (2016). When are readership counts as useful as citation counts? S copus versus M endeley for LIS journals. *Journal of the Association for Information Science and Technology*, 67(1), 191-199.

Meschede, C. & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. Scientometrics, 115(1), 283-297.

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.

Nielsen, J. (1997). Zipf curves and website popularity. *Nielsen Norman group*. Available from: http://www.nngroup.com/articles/zipf-curves-and-website-popularity/

Orduna-Malea, E., Martín-Martín, A., Thelwall, M., & López-Cózar, E. D. (2017). Do ResearchGate Scores create ghost academic reputations?. *Scientometrics*, *112*(1), 443-460.

Ortega, J. L. (2018a), Disciplinary differences of the impact of altmetric, *FEMS Microbiology Letters*, 365(7): fny049

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

Ortega, J. L. (2019a), Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact, *Aslib Journal of Information Management*,71(1): 124-136

Ortega, J. L. (2019b). The coverage of blogs and news in the three major altmetric data providers. In: *17th International Conference of the International Society for Scientometrics and Informetrics, Rome, Italy*

Robinson-Garcia, N., Arroyo-Machado, W., Torres-Salinas, D. (2019). Mapping social media attention in Microbiology: Identifying main topics and actors. *FEMS microbiology letters*, 366(7): fnz075. https://doi.org/10.1093/femsle%2Ffnz075

Ruan, Q. Z., Chen, A. D., Cohen, J. B., Singhal, D., Lin, S. J., & Lee, B. T. (2018). Alternative Metrics of Scholarly Output: The Relationship among Altmetric Score, Mendeley Reader Score, Citations, and Downloads in Plastic and Reconstructive Surgery. *Plastic and reconstructive surgery*, 141(3), 801-809.

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.

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11-21.

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

Thelwall, M. (2017a). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa?. *Journal of Librarianship and Information Science*, *49*(2), 144-151.

Thelwall, M. (2017b). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, *11*(1), 128-151.

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 1-10.

Wang, X., Liu, C., Fang, Z., & Mao, W. (2014). From Attention to Citation, What and How Does Altmetrics Work?. *arXiv preprint arXiv:1409.4269*.

Yu, H., Cao, X., Xiao, T., & Yang, Z. (2019). Accuracy of Policy Document Mentions: The Role of Altmetrics Databases. In: In *17th International Conference on Scientometrics and Informetrics, ISSI 2019* (pp. 477-488). Rome: International Society for Scientometrics and Informetrics.

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.

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.

Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, *68*(10), 2511-2521.

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort. Cambridge, Massachusetts: Addison-Wesley. p. 1.