Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact

José Luis Ortega

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

Abstract

Purpose: This study aims to analyse the metrics provided by Publons about the scoring of publications and their relationship with impact measurements (bibliometric and altmetric indicators).

Design/methodology/approach: In January 2018, 45,819 research articles were extracted from Publons, including all their metrics (scores, number of pre and post reviews, reviewers, etc.). Using the DOI identifier, other metrics from altmetric providers were gathered to compare the scores of those publications in Publons with their bibliometric and altmetric impact in PlumX, Altmetric.com and Crossref Event Data (CED).

Findings: The results show that (1) there are important biases in the coverage of Publons according to disciplines and publishers; (2) metrics from Publons present several problems as research evaluation indicators; and (3) correlations between bibliometric and altmetric counts and the Publons metrics are very weak (r<.2) and not significant.

Originality/value: This is the first study about the Publons metrics at article level and their relationship with other quantitative measures such as bibliometric and altmetric indicators.

Keywords: Publons, Altmetrics, Bibliometrics, Peer-review

1. Introduction

Traditionally, peer-review has been the most appropriate way to validate scientific advances. Since the first beginning of the scientific revolution, scientific theories and discoveries were discussed and agreed by the research community, as a way to confirm and accept new knowledge. This validation process has arrived until our days as a suitable tool for accepting the most relevant manuscripts to academic journals, allocating research funds or selecting and promoting scientific staff. However, this system presents two important limitations: expensive and subjective. Peer-review requires the involvement of two or more scholars that study and analyse each research unit (publication, institution, researcher, etc.) and then present an assessment report. This process consumes large amount of economic and time resources. Equally, peer-review suffers from subjective judgements and it would cause arbitrary and biased decisions that undermine the evaluation system.

Likewise, the professionalization of science in the nineteenth century (Beer and Lewis, 1963) caused a rapid increase of economic and human resources and, in consequence, an exponential growth of scholarly publications (Price, 1961). Bibliometrics indicators emerged as a complementary way, less expensive and more objective, of assessing complex academic scenarios resulted from this growing professionalization. Based on production and impact indicators, bibliometrics contributes indicators that allow to benchmark and assess the performance of different research units into disciplinary or institutional environments. In the manner of peer-review, bibliometrics also has important limitations such as manipulation and misuse (Narin et al., 1994). Practices such as salami publishing, abusive self-citations or using journal metrics to evaluate articles or authors question the suitability of bibliometrics for research evaluation.

Recently, the appearance of online social networks and web platforms for scientists is causing the proliferation of new metrics that quantify the use and impact of research outputs on the Web. Altmetrics introduce a new view on the importance of the dissemination and the social impact in research evaluation (Holmberg, 2015). In this context, several platforms (Publons, Faculty of 1000) are providing metrics that score publications according to user opinions. Publons is a web platform that allows their members to make public their outputs as journal reviewers. In addition, they can rate the quality of the reviewed articles, scoring their quality and relevance. This procedure suggests a new way of open peer-reviewing, in which the assessment of manuscripts are done in a transparent and public environment.

The aim of this study, therefore, is to explore Publons (publons.com), analysing its coverage and studying the relationship between its metrics and other bibliometric and altmetric indicators, with the aim of observing to what extent these scores can be associated to research impact.

2. Literature review

Since the strengthening of bibliometrics as research discipline and its employment in research evaluation processes, numerous studies have explore the relationship between these quantitative metrics and the result of the peer-review. Moed et al. (1985) published one of the first studies that compared bibliometric indicators with the result of peer judgments. The analysis revealed a serious lack of agreement between the two methods. However, Nederhof and Van Raan (1993) studied trends in productivity and impact of six economics research groups and compared their results with a peer-review study. They concluded that results from peer review and bibliometric studies appear to be complementary and mutually supportive. Thomas and Watkins (1998) found high correlations between peer-review and citations-based rankings of academic journals in a specific research discipline. More recently, Opthof et al. (2002) observed that reviewers' recommendations and editor's ratings were positively correlated with citations when they analysed the submissions to *Cardiovascular Research* between 1997 and 2002. Aksnes and Taxt (2004) analysed research groups at the University of Bergen, investigating the relationship between bibliometric indicators and the outcomes of peer-review. Their findings showed positive but relatively weak correlations. In the same vein, Patterson and Harris (2009) found a low but statistically significant correlation between citations and quality scores in papers published in the journal *Physics in Medicine and Biology*.

Finally, Van Raan (2006) provided comparable results between the h-index and the scores given to 147 university chemistry research groups by a review panel. His results showed that the h-index and bibliometric indicators relate in a similar way with peer judgments.

The appearance of web platforms that include opinions and scores of research papers (for example, Publons and F1000Prime of Faculty of 1000) is providing a new opportunity to match bibliometrics indicators, altmetrics and scores of researchers to assess the quality of scholarly outputs. Faculty of 1000 (F1000) was the first platform (2000) that enabled the valuation of published articles using a scoring system. Wardle (2010), who studied the relationship between citations and F1000 recommendations, published the first study on this system. He concluded that F1000Prime metrics cannot identify those publications that subsequently have the greatest impact. Li and Thelwall (2012) compared the scores of more than one thousand research papers from F1000Prime with the number of readers in Mendeley and bibliometric indicators. The moderate correlations (r=<.4) suggested that F1000Prime metrics measure different perspectives of research. Bornmann and Leydesdorff (2013) used F1000 recommendations to validate bibliometric indicators from InCites and they found that Percentile in Subject Area achieves the highest correlation with F1000 ratings. Waltman and Costas (2014) also found weak correlations between F1000 recommendations and citations. They interpreted this result to mean that F1000 fails to identify the most relevant articles or both measures do not capture the same type of impact. According to the relationship between altmetric and bibliometric indicators, most of the studies conclude that there is little relationship between alternative metrics and biblometrics (Priem et al., 2012; Thelwall et al., 2013; Costas et al., 2014), which could mean that they express a different type of impact.

Publons, however, has attracted the attention of few studies and many of them are merely descriptive analyses about its functionalities (Meyts et al., 2016; Sammour, 2016). This is because Publons is more a site for helping reviewers gain credit for their work than a service to assess publications. On this matter, we can highlight the study of Ortega (2017) who compared the peer-review activity and the bibliometric performance of Publons' users, finding weak correlations between both scholarly activities. Meadows (2017a, 2017b) analysed the ORCID profiles associated to peer-review platforms and she found that Publons is the top site with 92% of ORCID users. Nevertheless, no study has analysed the coverage of Publons, their article metrics and how they are related to other impact measurements.

Altmetric providers are becoming important tools to obtain and analyse altmetric and bibliometric data. Many studies have explored the reliability of data providers, analysing the coverage of publications and events. Jobmann et al. (2014) were the first ones to compare ImpactStory, Altmetric.com, PlumX and Webometric Analyst. They found that PlumX is the platform that better covers Mendeley and Facebook data, while Altmetric.com stands out gathering blogs, news and CiteULike data. Zahedi et al. (2015) explored the consistency of Altmetric.com, Mendeley and Lagotto. They also detected significant differences, finding that Altmetric.com gathers more tweets, but it is less accurate collecting Mendeley readers. More recently, Meschede and Siebenlist (2018) found that less than half of the publications analyzed are included in Altmetric.com, while PlumX covers almost the totality (99%). Ortega (2018a) compared several data providers and he found that Altmetric.com is better covering social metrics, PlumX gathering Mendeley readers and CED capturing Wikipedia citations. Zahedi and

Costas (2018) performed the most exhaustive comparison between data providers, resulting that the use of one or another altmetric provider has important effects on the results.

3. Objectives

The main objective of this study is to explore Publons as web service specialized in the sharing of peer-reviews reports and the publication of open post-publication reviews. This service provides the opportunity of analysing the relationship between the qualitative opinion of scholars about research publications and its connection with the bibliometric and altmetric impact. In other words, this study attempts to determine whether bibliometric and altmetric indicators can be associated to individual evaluations. Three research questions were formulated:

- How are the publications covered by Publons distributed? Is it possible to identify any bias that brings into question the reliability of Publons as research evaluation tool?
- How are the Publons metrics distributed? Could one find any limitation or weakness as assessment metrics?
- Is there any relationship between bibliometric and altmetric counts and the Publons metrics? Could research impact (bibliometric and altmetric indicators) be associated to subjective valuations (Publons' indicators)?

4. Methods

This study has used several sources to extract and gather the data. This is because several studies have evidenced that some providers cover some metrics better than others (Jobmann et al., 2014; Zahedi et al., 2015; Ortega, 2018a). In addition, the study selects only the metrics that have more incidence and higher values in each data provider (Ortega, 2018b):

Altmetric.com (altmetric.com): It was the first altmetric provider and was born in 2012 by Euan Adie, with the support of Digital Science. Altmetric.com is centred 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 (Application Programming Interface). Today, Altmetric.com tracks the social impact of close to 9 million of research papers. However, this platform does not include metrics about citations and usage. Most of the metrics were selected from this provider (Blogs, Facebook pages, Google+ users, News outlets, Reddit posts, Tweeters, CiteULike saves). The remaining metrics (Weibo, Youtube, Linkedin, Peer_review, Pinterest, Policy_papers, Questions, and Research_highlights) were discarded due to their low incidence.

PlumX (plumanalytics.com): PlumX is a provider of alternative metrics created in 2012 by Andrea Michalek and Michael Buschman. 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 artefacts, being then the largest altmetric aggregator. In 2017, Plum Analytics was acquired by Elsevier, allowing now to know the altmetric information of any document indexed in Scopus. Metrics about usage (Linkouts, Abstract views and Html views) and Mendeley readers were extracted from this source. Just as Altmetric.com, the remaining

metrics were dismissed due to their low importance (EPrints_downloads, PDF_views, SSRN_download, etc.).

Crossref Event data (www.crossref.org/services/event-data/) (CED): CED is the youngest service, created in 2016 and officially released in 2017. Due to this, the platform claims that the service is still in beta. Unlike the previous ones, CED is not a commercial site and it provides free access to their data though a public API. Another difference is that it does not aggregate the information, but it displays the entire information about each altmetric event. For instead, 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, data have to be processed to be comparable with the other services. CED was used exclusively to obtain Wikipedia citations.

Publons: Publons is a web platform created by Andrew Preston and Daniel Johnston in New Zealand in 2013, and acquired by Clarivate Analytics in 2017. The service is addressed to the scholarly community and its purpose is to create an open space that may improve the peer-review system, making it faster, more efficient and effective. Publons lists the reviewed publications besides to some metrics that value the quality of papers. These scores are assigned by the members when they upload the review to Publons and the result is made public only when the article is already published (Publons, 2018). Metrics used are:

- **Quality**: from 1 to 10, reviewers value the quality of the publication according whether the research has been well executed and designed, if the methods are sufficiently explained to be reproduced and if the conclusions are supported by the data. When the publication has been scored more than once, the result is the average.
- **Significance**: from 1 to 10, Significance measures the relevance and novelty of the publication, evaluating whether the article offers new insight into the field, if it could encourage new research lines and if it could be interesting to a wider audience. When the publication has been scored more than once, the result is the average.
- **Overall Publons score**: This score is the average between Quality and Significance, and it summarizes the quality of a paper.
- **Number of reviews**: the number of reviews received by an article. It could be several reviews from the same reviewer and these may be done pre-publication or post-publication.
- **Scores**: the number of individual opinions about an article. Contrary to Number of reviews, each member may only score an article once.
- **WoS citations**: the number of citations that an article receives. This metric comes from the Web of Science (WoS) because it also belongs to Clarivate.

4.1. Data extraction

The first step was to retrieve the most exhaustive sample of research articles from Publons. In Publications (https://publons.com/publon/?order_by=date), the complete list of articles by Research field were retrieved and scraped for relevant data. Then, the site was crawled using the Publons identifier and extracting the metrics and bibliographic information about each

paper. Finally, 45,819 articles were obtained from Publons. This process was carried out during January 2018.

From these articles, DOI identifier was obtained and used to retrieve altmetric information from the three Altmetric providers: Altmetric.com, PlumX and CED. The coverage of these tools was dissimilar. PlumX is the service with the best coverage, indexing 45,281 articles (98.8%); Altmetric.com gathers 28,807 (63.6%), a third less than PlumX; and CED captures only 1,583 articles mentioned on Wikipedia (3.5%). This process was done during February 2018.

4.2. Altmetrics

Indicator	Metrics	Sources	Provider	Definition
Downloads	Downloads	Airiti Library, bepress, Dryad, DSpace, ePrints, Figshare, Github, Institutional Repositories, Pure, RePEc, Slideshare, SSRN	PlumX	The number of times a publication is downloaded from different platforms.
	Abstract views	bit.ly	PlumX	The number of times the abstract of an article has been viewed
	Link Outs	EBSCO databases	PlumX	The number of times an article's URL is clicked
Views	HTML views	Airiti Library, bepress, CABI, DSpace, EBSCO, ePrints, PLOS, RePEc, SSRN	PlumX	The number of times an article has been viewed or clicked
	Clicks		PlumX	
	PDF views		PlumX	
Readers		Mendeley	PlumX, Altmetric.com	The number of saves of a document into a user's library
Citations		Scopus, WoS	PlumX	The number of times that a paper is cited by other publications
Tweets		Gnip (Twitter)	PlumX, Altmetric.com	The number of tweets and retweets that mention a research paper
Blog mentions		PlumX: 55,000 media and blogs (Newsflo) Altmetric.com:	PlumX, Altmetric.com	The number of blog posts written about one article

		11,000 blogs		
Google+		Google+	Altmetric.com	The number of public posts about one article
News		PlumX: 55,000 media and blogs (Newsflo) Altmetric.com: 2,700 media	PlumX, Altmetric.com	The number of news outlets that mention one article
Reddit		Reddit	Altmetric.com	The number of original posts about one article
CiteULike		CiteUlike	Altmetric.com	The number of bookmarks of a document
Comments		Facebook	Altmetric.com, PlumX	The number of posts on public Facebook pages about one article
Wikipedia		Wikipedia	Altmetric.com, PlumX, Crossref Event Data	The number of Wikipedia entries that cite one article

Table 1. Metrics, sources and definition of the altmetric indicators used in the study

Table 1 details the list of altmetrics used in this study besides their definition, the source and providers of each metric. When several providers supply the same metric (News, Blogs, etc.), the highest count in each source was selected. For example, whether one article has one blog mention in PlumX and two in Altmetric.com, the two mentions of Altmetric.com were used. In this way, it is attempted to present the most exhaustive picture of the altmetric impact in each publication.

5. Results

5.1. Coverage

This section shows the distribution of reviewed articles in Publons by research area and publisher.

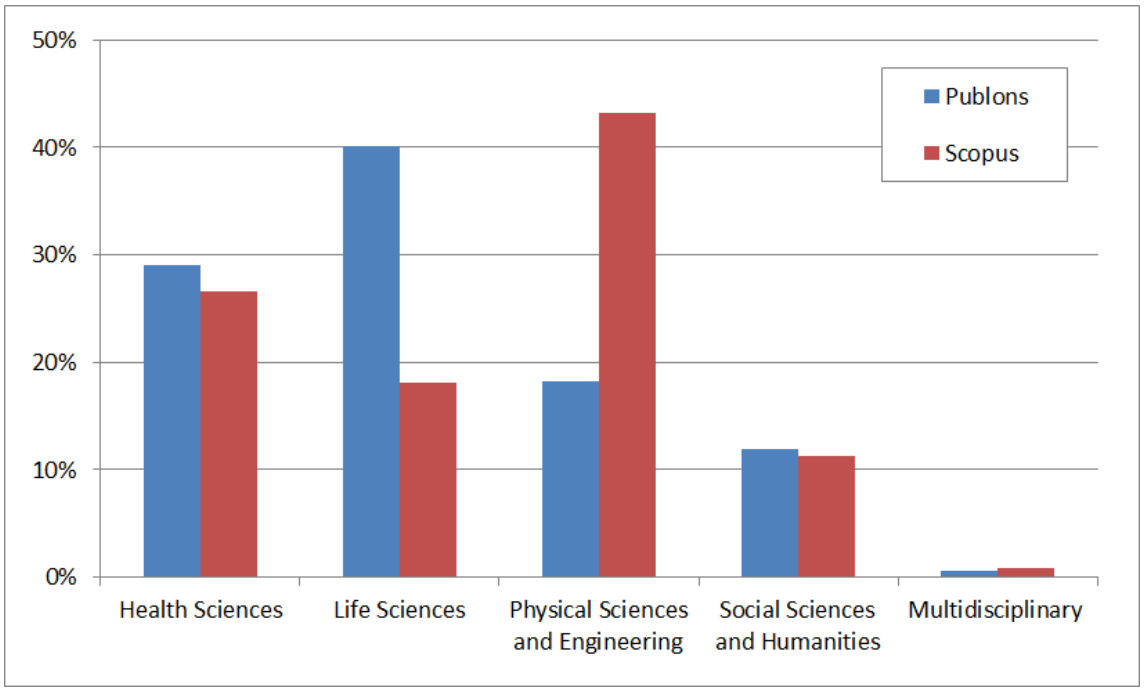

Figure 1. Percentage of articles in Publons and Scopus by Research Area.

Figure 1 displays the proportion of articles in Publons and Scopus by research area. This comparison was made with Scopus because this is one of the largest multidisciplinary citation indexes with a well-balanced distribution of publications (Miguel et al., 2011; Mongeon and Paul-Hus, 2016). Another reason is that both sites share the same subject classification (All Science Journal Classification, ASJC). Scopus data were retrieved from its search page. The bar plot clearly shows that Publons is biased to *Life Sciences* (40.2%), with more than the double of publications than Scopus (18.3%). By contrast, the number of publications from *Physical Sciences* and *Engineering* is much lower in Publons (18.3%) than in Scopus database (43.2%). At the level of the Subjects Areas, *Neurosciences* (Publons=11.3%; Scopus=1.6%), *Psychology* (Publons=7%; Scopus=1.4%) and *Immunology and Microbiology* (Publons=5.7%; Scopus=1.8%) are the most overrepresented disciplines in Publons. Conversely, *Engineering* (Publons=5%; Scopus=11.2%), *Physics and Astronomy* (Publons=3%; Scopus=7%) and *Social Sciences* (Publons =1.6%; Scopus=4.6%) have lower percentages of publications in Publons than in Scopus.

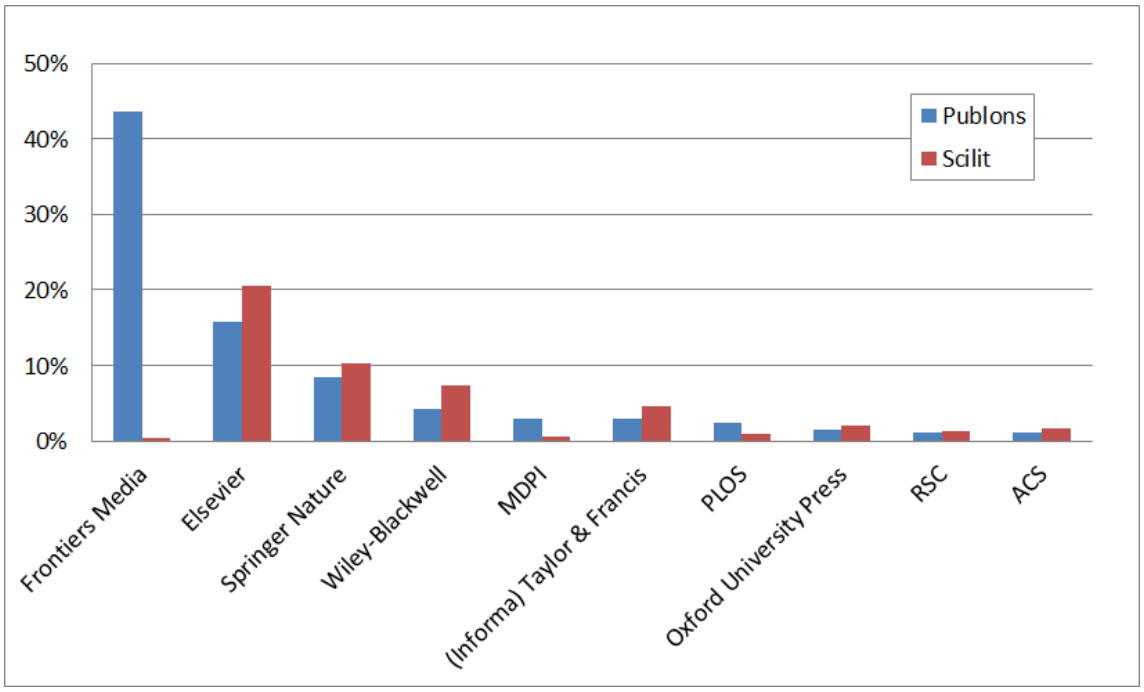

Figure 2. Percentage of reviewed articles by publisher in Publons and percentage of published papers by publisher in Scilit.

Publons also displays important differences in the coverage of articles according to publishing house. Publons has signed several agreements with publishers to automatically upload the result of reviews to the platform (Publons, 2017). This fact would introduce some possible biases because it favours the uploading of reviews from specific journals. Figure 2 shows the percentage of articles by publisher in Publons and Scilit (www.scilit.net), a website that provides the most comprehensive figures on scientific literature. Scilit includes publication data from 2012 to 2017. The most overrepresented publishing groups in Publons are *Frontiers Media* (Publons=44%, Scilit=.5%), *MDPI* (Multidisciplinary Digital Publishing Institute) (Publons=29%, Scilit=.6%) and *PLOS* (Public Library of Science) (Publons=23%, Scilit=1%). Curiously enough, these are the main open access publishers which it suggests that the firms that promote open access are also involved in making public their review results. The case of *Frontiers Media* is among the most significant because the greatest part of the articles published by this publisher are uploaded to Publons which could be due to any type of special agreement between both platforms. However, the pilot partnership with SAGE has not been reflected in the study, where the number of reviewed articles from SAGE does not reach 1% (Research Information, 2015) in relation to the 2% of published papers in Scilit.

5.2. Distribution of Publons metrics

Next, statistical distributions of the metrics provided by Publons are analysed. The objective is to describe the prevalence and meaning of these metrics in the total number of articles included in Publons.

	Pre-pub reviews	Pre-pub reviews %	Post-pub reviews	Post-pub reviews %	Scores	Scores %
Total articles	38,447	83.5%	2,238	4.9%	25,096	54.5%

1 reviewer	14,482	37.7%	2,087	93.3%	23,159	92.3%
2 reviewers	18,976	49.4%	108	4.8%	1,825	7.3%
3 reviewers	3,887	10.1%	23	1.0%	107	0.4%
>3 reviewers	1,100	2.9%	20	0.9%	5	0.0%
Mean	1.797		1.099		1.082	

Table 2. Parameters of statistical distributions of Publons' production metrics

Table 2 shows the distribution of Publons' production metrics, that is, metrics about the number of reviews and scores. Overall, most of the articles are reviewed before publication (83.5%), while a much-reduced number of papers are also reviewed after publication (4.9%). These differences inform us that Publons users are using the platform to upload pre-publication reviews more than to work as reviewers into the platform. This assumption is reinforced by the fact that just over half of papers have been scored into Publons (54.5%). According to the number of reviews, pre-publication manuscripts are reviewed once (37.7%) or twice (49.4%). Meanwhile, most of the post-publications reviews are done once (93.3%). The same happen with Scores, which are mostly reviewed by only one reviewer (92.3%). This last figure raises some questions about the reliability of this metric because it is mostly based on the opinion of only one person.

	Publons score	**Significance**	**Quality**
Mean	7.168	7.132	7.166
N	25,096	25,096	25,096
Standard Deviation	1.793	1.911	1.842
1th Quartile	6.000	6.000	6.000
Median	7.500	7.000	7.000
3rd Quartile	8.500	8.000	8.000

Table 3. Descriptive statistics about Publons score and the variables of that metric.

Table 3 displays several parameters of Publons score and the two variables that compose it, Significance and Quality. The first interesting thing is that the means of the three metrics are relatively high, 7.17 for Publons score, 7.13 for Significance and 7.17 for Quality. According to the quartiles, the 75% of the articles have a score higher than 6 and 25% of them reach more than 8.5 points. This fact suggests that users tend to positively score the uploaded articles, avoiding negative or critical assessments. The small Standard Deviation (SD<2) also suggests that the scores are rather similar. Another interesting point is the small difference between Significance and Quality. Both metrics provide very similar means and quartiles. These important similarities suggest that users do not distinguish the meaning of both metrics or, perhaps, they think that quality is strongly associated with significance.

5.3. Correlations

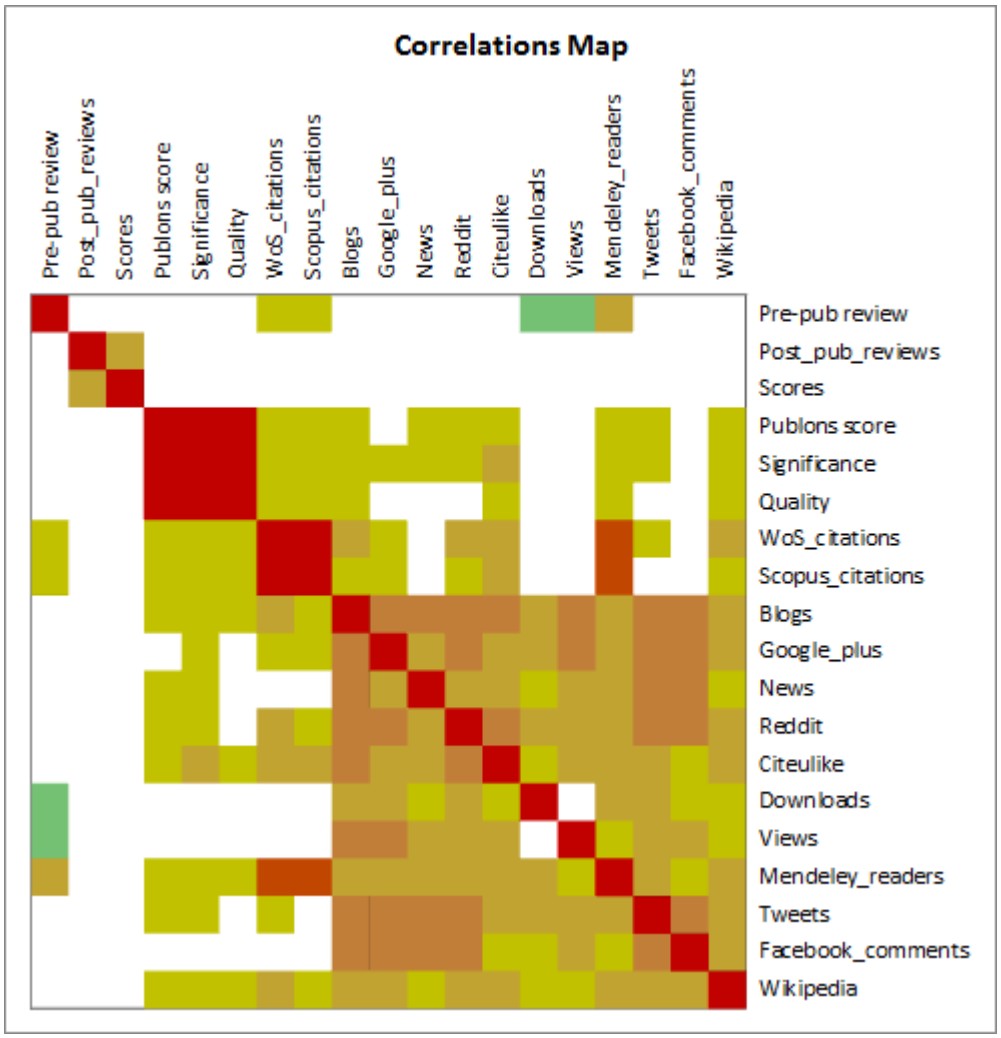

Figure 3. Correlation map between altmetrics, bibliometrics and Publons metrics.

Figure 3 shows the Pearson correlation map between altmetrics, bibliometrics and Publons metrics. Overall, the map shows low correlations among all the variables, with a correlation average of *r*=.22. This result suggests that there is little association between these different types of metrics. The only exceptions are found in the high correlations between Publons score and the variables that comprise that indicator, Significance (*r*=.954) and Quality (*r*=.959). It is interesting that both metrics have a high correlation between them (*r*=.84); in spite of they should measure different aspects of an article. This strong correlation confirms the previous results about the high similarity between Significance and Quality. There is also high correlation between citation indexes (WoS and Scopus) (*r*=.96) due to citations are similarly distributed in both bibliographic databases (Archambault et al., 2009). Finally, Mendeley readers shows an important correlation with citations (r-Scopus=.68; r-WoS=.66), which confirms that this altmetric is the best associated with bibliometric impact (Mohammadi et al, 2015; Maflahi and Thelwall, 2016; Thelwall, 2017).

The remaining metrics show very low correlations, especially between the altmetrics and Publons metrics. In these cases, correlations do not exceed .2. For instance, the number of reviews and scores do not show any significant relationship with any variable, and Publons score, Significance and Quality show very poor correlations with the other metrics (*r*<.2). The

picture also shows that altmetrics better correlate among them than with Publons metrics. The result of these correlations suggests therefore that there is no relationship between the quality expressed in Publons metrics and the impact measured by bibliometrics and altmetrics.

6. Discussion

The exploratory study of Publons has made possible to analyse the distribution of reviewed publications by discipline and publisher. The results show that Publons is biased in favour to *Health Sciences* and *Life Sciences*, while it shows an important gap of *Physical Sciences* and *Engineering* reviews. This distribution is slightly different to Meadows (2017b) and it could be due to she only uses reviews from ORCID profiles. These differences would suggest that researchers from *Health Sciences* and *Life Sciences* areas could be more interested in publishing their reviews, whereas physicians and engineers are reluctant to share their review reports. These disciplinary differences were also observed by Burley (2017) when she verified that open peer review in Nature Communications is adopted more frequently by life sciences authors than chemists and physicist. Ross-Hellauer et al. (2017) also perceived that researchers from applied sciences were more reserved towards opening up peer their review reports. The causes of these different attitudes might be due to the role of the peer review in each research area (Weller, 2001). From a research evaluation point of view, this disciplinary bias could distort the assessment of articles from underrepresented disciplines when they are compared to the total. Another important limitation is the biased coverage of publishers. The results have showed that Publons includes more articles from open access platforms, especially from *Frontiers Media*. These differences could favour the assessment of publications from open access publishers to the detriment of articles from other publishers.

With regard to the representativeness of Publons metrics as quality indicators, the results cast some doubts about them. The first problem is that one user only scores most of the articles (92%). Thus, the quality of many of the articles is based only on the subjective opinion of one specific reviewer. The absence of different opinions, which reach a minimal agreement, make that all the scores could be at the mercy of interested and conflictive statements (i.e. friendship, enmity, etc.) (Peters and Ceci, 1982; Travis and Collins, 1991). This problem is even more serious because it is demonstrated that the opinion of reviewers about a paper is very different (Rothwell and Martyn, 2000; Neff and Olden, 2006; Kravitz et al., 2010). Another problem, and perhaps as consequence of the previous limitation, is the similar distribution of Significance and Quality, the two components of the Publons score. These metrics should have different meanings because they measure the impact and importance of papers (Significance) and their methodological rigor (Quality). However, the absence of disparity suggests that users interpret these variables in the same way.

The correlations analysis has evidenced that there is no relationship between the metrics provided by Publons and the altmetrics and bibliometrics from PlumX, Altmetric.com and CED (r<.2). The low values for the correlation coefficients demonstrate that expert scores and the number of reviews are not associated with the altmetric and bibliometric impact. This result could mean two things. First, it is possible that Publons score could not be a suitable quality indicator and the lack of high and positive correlations would demonstrate that Publons score, due mainly to the aforementioned problems, does not measure the quality and impact of a

paper properly. In this sense, Waltman and Costas (2014) came to a similar conclusion when they studied the relationship between F1000 recommendations and citations. The weak correlation between both measures suggested that F1000 fails to identify the most important publications. In contrast, another explanation could be that bibliometric and altmetric indicators could not be entirely associated with quality, but with the social or academic impact. That is, the number of mentions and citations would be strongly influenced by the way in which an article is disseminated (choosing a journal, being open or not open access, sharing in academic sites, promoting in social networks, etc.). This perspective assumes that the spreading could be more important than the quality in bibliometric and altmetric indicators. In fact, many authors have doubted that bibliometric indicators can describe quality aspects of a publication (Lindsey, 1989; Seglen, 1997; Nieminen et al., 2006). However, the positive association between peer-review assessments and bibliometric indicators (Thomas and Watkins, 1998; Opthof et al., 2002; Van Raan, 2006; Patterson and Harris, 2009) suggest that citations could reflect, to a great extent, the quality of an article. Therefore, the lack of positive and significant correlations between Publons metrics and bibliometric and altmetric indicators would be caused by the inconsistencies and limitations in the design of the measures proposed by the peer-review platform.

The main limitation of this study is the inconsistent coverage of the altmetric providers that could not show the entire altmetric impact of a document. The low coverage of Altmetric.com (63.6%) and the limitations expressed by other studies (Meschede and Siebenlist, 2018; Ortega, 2018b; Zahedi and Costas, 2018), could present a non-realistic picture of the relationship between Publons scores and other metrics. However, in our opinion, to solve this limitation would not improve the poor correlation between these measures.

7. Conclusions

The results show that there are important biases in the coverage of Publons according to disciplines and publishers. From a disciplinary point of view, Publons is biased in favour of *Health Sciences* and *Life Sciences*, while *Physical Sciences* and *Engineering* are underrepresented. With regard to publishers, Publons includes more articles from open access platforms. These biases could be significant when it comes to performing scientometrics analysis.

Metrics from Publons present several problems as research evaluation indicators. The results show that most of the scores are based in the particular opinion of one user. In addition, users do not distinguish between Quality and Significance, being scored in the same way. These limitations cast doubts about the reliability of Publon score to be used as quality indicator.

Correlations between bibliometric and altmetric counts and the Publons metrics are very weak and not significant. This fact evidences that both type of metrics are not related among them. In my opinion, this lack of relation could be due to inconsistences and limitations in the design of Publons metrics, which do not allow it to capture the opinion of the users in a proper way.

8. References

Aksnes, D. W. and Taxt, R. E. (2004), “Peer reviews and bibliometric indicators: a comparative study at a Norwegian university”, *Research evaluation*, Vol. 13 No. 1, pp. 33-41.

Archambault, É., Campbell, D., Gingras, Y. and Larivière, V. (2009), “Comparing bibliometric statistics obtained from the Web of Science and Scopus”, *Journal of the Association for Information Science and Technology*, Vol. 60 No. 7, pp. 1320-1326.

Beer, J. J. and Lewis, W. D. (1963), “Aspects of the Professionalization of Science”, *Daedalus*, Vol. 92 No. 4, pp. 764-784.

Bornmann, L. and Leydesdorff, L. (2013), “The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000”, *Journal of Informetrics,* Vol. 7 No. 2*,* pp. 286–291*.*

Burley, R. (2017), “Peer review in the 21st century”, *Information Services & Use*, Vol. 37 No. 3, pp. 259–261.

Costas, R., Zahedi, Z. and 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*, Vol. 66 No. 10, pp. 2003-2019.

Holmberg, K. J. (2015), *Altmetrics for information professionals: Past, present and future*. Cambridge, UK: Chandos Publishing.

Jobmann, A., Hoffmann, C.P., Künne, S., Peters, I., Schmitz, J. and Wollnik-Korn, G. (2014), “Altmetrics for large, multidisciplinary research groups: Comparison of current tools”, *Bibliometrie – Praxis und Forschung*, No. 3, http://www.bibliometrie-pf.de/article/viewFile/205/258

Kravitz, R. L., Franks, P., Feldman, M. D., Gerrity, M., Byrne, C. and Tierney, W. M. (2010), “Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care?”, *PLoS One*, Vol. 5 No.4, pp. e10072.

Li, X. and 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). OST and Science-Metrix.

Lindsey, D. (1989), “Using citation counts as a measure of quality in science measuring what's measurable rather than what's valid”, *Scientometrics*, Vol. 15 No. 3-4, pp. 189-203.

Maflahi, N. and Thelwall, M. (2016), “When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals”, *Journal of the Association for Information Science and Technology*, Vol. 67 No.1, pp. 191-199.

Meadows, A. (2017a), “The Role of Persistent Identifiers in the Peer Review Process: Use of ORCID”, in *Eighth International Congress on Peer Review and Scientific Publication*. September 10-12, 2017 Chicago, Illinois, USA https://peerreviewcongress.org/prc17-0366

Meadows, A. (2017b), "Peer Review at ORCID - An Update", *ORCID blog*. https://orcid.org/blog/2017/09/14/peer-review-orcid-update

Meschede, C. and Siebenlist, T. (2018), "Cross-metric compatability and inconsistencies of altmetrics", *Scientometrics*, Vol. 115 No. 1, pp. 283-297.

Meyts, R. D., Losito, S. and Carrell, D. T. (2016), "Rewarding peer-review work: the Publons initiative", *Andrology*, Vol. 4 No. 6, pp. 985-986.

Miguel, S., Chinchilla-Rodriguez, Z. and de Moya-Anegón, F. (2011), "Open access and Scopus: A new approach to scientific visibility from the standpoint of access", *Journal of the Association for Information Science and Technology*, Vol. 62 No. 6, pp. 1130-1145.

Moed, H. F., Burger, W. J. M., Frankfort, J. G. and Van Raan, A. F. (1985), "A comparative study of bibliometric past performance analysis and peer judgement", *Scientometrics*, Vol. 8 No 3-4, pp. 149-159.

Mohammadi, E., Thelwall, M., Haustein, S. and Larivière, V. (2015), "Who reads research articles? An altmetrics analysis of Mendeley user categories", Journal of the Association for Information Science and Technology, Vol. 66 No. 9, pp. 1832-1846.

Mongeon, P. and Paul-Hus, A. (2016), "The journal coverage of Web of Science and Scopus: a comparative analysis", *Scientometrics*, Vol. 106 No. 1, pp. 213-228.

Narin, F., Olivastro, D. and Stevens, K. A. (1994), "Bibliometrics/theory, practice and problems", *Evaluation review*, Vol. 18 No.1, pp. 65-76.

Neff, B. D. and Olden, J. D. (2006), "Is peer review a game of chance?", *AIBS Bulletin*, Vol. 56 No. 4, pp. 333-340.

Nieminen, P., Carpenter, J., Rucker, G. and Schumacher, M. (2006), "The relationship between quality of research and citation frequency", *BMC Medical Research Methodology*, Vol. 6 No. 1, pp. 42.

Opthof, T., Coronel, R. and Janse, M. J. (2002), "The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias", *Cardiovascular research*, Vol. 56 No. 3, pp. 339-346.

Ortega, J. L. (2017), "Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons", *Scientometrics*, Vol. 112 No. 2, pp. 947-962.

Ortega, J. L. (2018a), "Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data", *Scientometrics*, Vol. 116 No. 3, pp. 2123-2138.

Ortega, J. L. (2018b), "Disciplinary differences of the impact of altmetric", *FEMS microbiology letters*, Vol. 365 No. 7, pp. fny049.

Patterson, M. and Harris, S. (2009), "The relationship between reviewers' quality-scores and number of citations for papers published in the journal Physics in Medicine and Biology from 2003–2005", *Scientometrics*, Vol. 80 No. 2, pp. 343-349.

Pautasso, M. and Schäfer, H. (2009), "Peer review delay and selectivity in ecology journals", *Scientometrics*, Vol. 84 No. 2, pp. 307-315.

Peters, D. P. and Ceci, S. J. (1982), "Peer-review practices of psychological journals: The fate of published articles, submitted again", *Behavioral and Brain Sciences*, Vol. 5 No. 2, pp. 187-195.

Price, D J de S. (1961), *Science since Babylon*. New Haven, CT: Yale University Press.

Priem, J., Piwowar, H. A. and Hemminger, B. M. (2012), "Altmetrics in the wild: Using social media to explore scholarly impact", arXiv preprint arXiv:1203.4745.

Publons (2017), "My journal is partnered with Publons. What does this mean?" https://publons.freshdesk.com/support/solutions/articles/12000012228-my-journal-is-partnered-with-publons-what-does-this-mean-

Publons (2018), "Publons Scoring". https://publons.freshdesk.com/support/solutions/articles/12000022210--publons-scoring

Research Information (2015), "Sage and Publons announce peer review pilot". https://www.researchinformation.info/news/sage-and-publons-announce-peer-review-pilot?news_id=1909

Ross-Hellauer, T., Deppe, A. and Schmidt, B. (2017), "Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers", *PloS one*, Vol. 12 No. 12, pp. e0189311.

Rothwell, P.M. and Martyn, C.N. (2000), "Reproducibility of peer review in clinical neuroscience: is agreement between reviewers any greater than would be expected by chance alone?", *Brain*, Vol. 123 No. 9, pp. 1964-1969.

Sammour, T. (2016), "Publons.com: credit where credit is due", *ANZ journal of surgery*, Vol. 86 No. 6, pp. 512-513.

Seglen, P. O. (1997), "Citations and journal impact factors: questionable indicators of research quality", *Allergy*, Vol. 52 No. 11, pp. 1050-1056.

Thelwall, M., Haustein, S., Larivière, V. and Sugimoto, C. R. (2013), "Do altmetrics work? Twitter and ten other social web services", PloS ONE, Vol. 8 No. 5, pp. 64841.

Thelwall, M. (2017), "Are Mendeley reader counts useful impact indicators in all fields?", *Scientometrics*, Vol. 113 No. 3, pp. 1721-1731.

Thomas, P. R. and Watkins, D. S. (1998), "Institutional research rankings via bibliometric analysis and direct peer review: A comparative case study with policy implications", *Scientometrics*, Vol. 41 No. 3, pp. 335-355.

Travis, G. D. L. and Collins, H. M. (1991), "New light on old boys: cognitive and institutional particularism in the peer review system", *Science, Technology & Human Values*, Vol. 16 No. 3, pp. 322-341.

Van Raan, A. F. (2006), "Comparison of the Hirsch-index with standard bibliometric

indicators and with peer judgment for 147 chemistry research groups", *Scientometrics*, Vol. 67 No. 3, pp. 491-502.

Waltman, L. and Costas, R. (2014), "F1000 recommendations as a potential new data source for research evaluation: A comparison with citations", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 3, pp. 433-445.

Wardle, D. A. (2010), "Do'Faculty of 1000'(F1000) ratings of ecological publications serve as reasonable predictors of their future impact?", *Ideas in Ecology and Evolution*, No. *3*, pp. 11-15.

Weller, A. C. (2001), *Editorial peer review: Its strengths and weaknesses*. Medford, NJ: Information Today, Inc.

Zahedi, Z., Fenner, M. and Costas, R. (2015), "Consistency among altmetrics data provider/aggregators: What are the challenges?", *The 2015 Altmetrics Workshop*, Amsterdam.

Zahedi, Z. and Costas, R. (2018), "General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators", *PloS one*, Vol. 13 No. 5, pp. e0197326.