Full paper

STI 2022 Conference Proceedings

Proceedings of the 26th International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Proceeding Editors

Nicolas Robinson-Garcia
Daniel Torres-Salinas
Wenceslao Arroyo-Machado

Citation: Torres-Salinas, D., Montero-Alonso, M. A., & Arroyo-Machado, W. (2022). Skewness distribution of four key altmetric indicators: an in-progress analysis across 22 fields in a national context. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti2232). https://doi.org/10.5281/zenodo.6960147

Collection: https://zenodo.org/communities/sti2022grx/

Skewness distribution of four key altmetric indicators: an in-progress analysis across 22 fields in a national context

Daniel Torres-Salinas[*], Miguel Ángel Montero-Alonso[**] and Wenceslao Arroyo-Machado[*]

[*] *torressalinas@ugr.es;* wences@ugr.es
Department of Information and Communication, University of Granada, Campus Universitario de Cartuja C.P. 18071, Granada, Spain

[**] *mmontero@ugr.es*
Statistics and Operational Research Department, University of Granada, Avenida de la Investigación nº 11 C.P. 18071, Granada, Spain

Introduction

The study of distributions in bibliometric data has been very common since Seglen's contributions in the 1990s (Seglen, 1992). However, the starting point was Lotka's seminal work in the 1920s, more specifically in 1926 (Lotka, 1926). Studies have shown that bibliometric distributions are asymmetric in terms of production, publications and number of citations. A limited number of items seem to account for most of the incidence. It has been established that the phenomenon of asymmetry occurs at paper level, individual level (Abramo et al., 2017) and unit level (Schmoch, 2020). Due to the asymmetry of bibliometric indicators, the use of measures such as the average has been questioned (Schmoch, 2020). Solutions to this problem include the removal of outliers or the use of normalised citation indicators (Radicchi et al., 2008).

Another aspect commonly studied is the asymmetry of different scientific fields, in an effort to establish whether it is the same in all scientific fields or whether there are certain fields where the asymmetry is more pronounced. For example, Albarrán and Ruiz-Castillo (2011) and Albarrán et al. (2011) studied 22 fields and 219 sub-fields respectively of the Web of Science, concluding that citations have highly skewed distributions. In the study by Schmoch (2020), 64% of the sub-fields had power laws and 2% of the publications accounted for 13.5% of all citations. Among the most recent studies, Bornmann and Leydesdorff (2017) studied six major scientific fields and the Impact Factor, assessing the skewness of citation distributions and the consequences of skewness in relation to covariates of citation scores. Finally, Waltman, Van Eck and Van Raan (2012) analysed 221 categories (science and social sciences) and verified the universality of citation distributions.

Skewness is therefore a widely and constantly studied phenomenon. In the more traditional bibliometric world, distributions, especially citation distributions, are the subject of in-depth analysis. However, this is not the case with altmetrics research, where the problem is more difficult to address. The large number of indicators (mentions, shares, reviews, comments, saves, etc.) and the variety of sources (Altmetric.com, PlumX, CrossRef Event, etc.) make this type of analysis more complex. However, there are some studies which have addressed

the analysis of altmetric distributions. The study by Zahedi et al. (2014) reported that the two most tweeted papers were from the field of Physics, receiving more than half of the total number of tweets for the year and thus showing a strong skewed distribution. Yu et al. (2017) concluded that mentions on Weibo and Twitter follow a distribution similar to 20-80.

Regarding other platforms, Costas et al. (2017) conducted a study on Mendeley in 30 different scientific fields. By comparing the number of citations and readership counts, they concluded that the distributions are highly skewed in all fields and their shapes are remarkably similar for both metrics. Thewall previously obtained similar results for Mendeley applied to different fields of Medicine (Thelwall & Wilson, 2016), noting the similarity between citations and readership counts: "*Spearman correlations were used to assess the strength of association between the citation counts and the readership counts because both data sets were skewed.*" Subsequent studies such as D'Angelo & Di Russ (2019) further confirm these results with Mendeley.

However, there is a lack of studies that comprehensively and systematically analyse asymmetry in the world of altmetrics. Studies are needed not only in terms of disciplines and specialties but also considering multiple altmetric indicators and platforms. Therefore, the aim of this article is to quantify the asymmetry in altmetrics, considering four indicators of mentions and 22 scientific areas. More precisely, the specific objectives are as follows:

- Sub-objective 1: to establish and compare the degree of asymmetry of four relevant altmetric indicators included in Altmetric.com (Twitter, Wikipedia, news and policy documents).
- Sub-objective 2: to determine whether the asymmetry of the four indicators is the same for all 22 Essential Science Indicators (ESI) science fields or whether there are areas where these indicators have different distributions.
- Sub-objective 3: to contextualise the results achieved by comparing them with the citation distributions, in this case the number of citations collected in the Dimensions database.

Material & Methods

The data were retrieved on 3 March 2021 from Web of Science, InCites and Altmetric.com. The Web of Science and InCites data consisted of 434,827 articles, editorial material, letters, and proceedings, papers and records published between 2016 and 2020 in which at least one of the authors had Spanish affiliation. The three main citation indexes and the Emerging Sources Citation Index (ESCI) were used. Multidisciplinary publications were reassigned to specific subject categories from Web of Science. For this purpose, InCites was used first, since it performs a category reassignment based on publication references, leaving 1,711 publications that had to be reassigned manually. Through the DOI, 237,232 of the publications were identified and retrieved from Altmetric.com. Finally, all these publications were classified into 22 general research fields, based on the 21 areas used by ESI plus Arts and Humanities. Although the ESI classification is at journal level, in this study it was applied at publication level. To do so, the 254 subject categories from Web of Science were matched with the ESI classification following the equivalence schema proposed by Tan (2020).

Table 1. Descriptive statistics of the different indicators used in the study

	News Mentions	Policy Mentions	Twitter Mentions	Wikipedia Mentions	Dimensions citations
Papers with mentions	26,167	4,569	208,253	5,989	202,833
Mean	0.846	0.033	13.42	0.047	14.127
Standard deviation	8.46	0.392	96.981	0.859	55.651
Maximum	1,429	58	15,695	208	6,573

One bibliometric indicator and four altmetric indicators were selected from the dataset (see Table 1). Two statistical measures were used to analyse the distributions of the indicators mentioned. The first and main one is the nonparametric skew, defined as being where the mean (μ), median (ν) and standard deviation (σ) of the population have their usual meanings. The nonparametric skew is a measure of the asymmetry of the probability distribution. Its value is zero and absolute values ≥ 0.2 indicate marked skewness. Nonparametric skew was chosen because other measures such as the kurtosis coefficient are not suitable for asymmetric distributions such as bibliometric distributions. Other authors (Opthof and Leydesdorff, 2010) have suggested that this kind of analysis may be performed using nonparametric statistics. The Gini Index was used as a complementary measure and to verify and compare the results, an indicator commonly used in these types of bibliometric studies (Bornmann & Leydesdorff, 2017). The Gini Index may be defined as a measure of inequality; a coefficient of 1 means that a single paper receives all citations and a coefficient of 0 means that the citations are equally distributed over the paper (Torres-Salinas et al., 2014).

Results

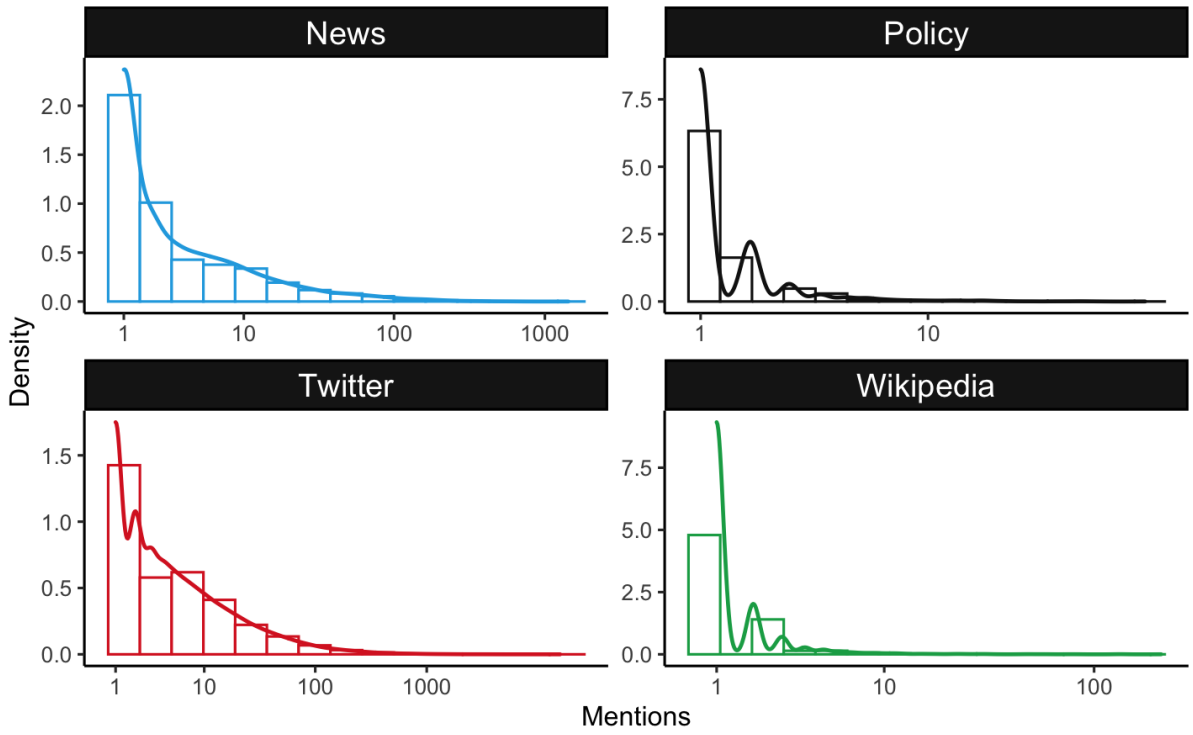
General indicators

Table 2. General indicators of asymmetry for four altmetric indicators

Statistical Indicator	News Mentions	Policy Mentions	Twitter Mentions	Wikipedia Mentions
Nonparametric skewness	0.097	0.085	0.104	0.055
Gini index	0.998	0.996	0.998	0.998

Figure 1 and Table 2 present the overall results of the altmetric indicators. The nonparametric skewness indicator and the Gini index present similar trends for the four indicators studied. If we consider the nonparametric skewness, the altmetric indicators with the lowest asymmetry are Wikipedia and policy document mentions, with values of 0.055 and 0.085 respectively. On the other hand, the altmetric indicator with the highest asymmetry scores is Twitter mentions, with a nonparametric skewness of 0.104 and a Gini index of 0.998. The Gini index (Table 2) does not seem particularly discriminating for three of the altmetrics since they have the same value (0.998). In any case, Figure 1 shows that the distributions of the four indicators follow two clear patterns. They may be grouped by similarity into two pairs: policy document mentions with Wikipedia mentions, and news mentions with Twitter mentions.

Figure 1. Histogram and distribution density of news, policy documents, Twitter and Wikipedia mentions with zeros excluded



Nonparametric skew applied to 22 Scientific Fields

Figure 2 and Table 3 show the differences of the nonparametric skew indicator in 22 research fields. The first aspect that may be observed is that each altmetric indicator has its own pattern. Twitter is the platform with the highest nonparametric skew values. It is particularly high in four research fields, where the threshold for high asymmetry (0.2) is exceeded. These are, from highest to lowest, the following scientific fields: "Pharmacology & Toxicology" (0.2386), "Chemistry" (0.2379), "Arts & Humanities" (0.2233) and "Plant & Animal Science" (0.2008). News mentions is the measure with the least variation, having a more homogeneous nonparametric skewness value across all fields. However, the fields with the lowest values may be highlighted: "Social Sciences" (0.0841), "Economics & Business" (0.0629) and "Mathematics" (0.0581). The highest value for news mentions was registered by "Space Sciences" (0.1648).

The case of policy mentions is singular. This altmetric indicator has the scientific fields with the lowest nonparametric skew. The fields with the lowest asymmetry are "Biology & Biochemistry" (0.0286), "Space Sciences" (0.0319) and "Materials Science Policy" (0.0297). One exception in the case of policy mentions is "Economics & Business" (0.1879). As for Wikipedia, "Arts & Humanities" has the highest value with 0.1838 and "Social Sciences, General" the lowest with 0.0399. Another interesting aspect is that if we look at a single scientific area, it may have a very different pattern for each of the four indicators. For example, "Chemistry" has strong asymmetry and one of the highest values (0.2379) in the Twitter mentions indicator. However, in the policy mentions indicator it has one of the lowest nonparametric skew values (0.0412). Such extreme patterns can be clearly observed in other scientific fields such as "Pharmacology & Toxicology" and "Arts & Humanities". In the case of the Gini index, the results and patterns are very similar to those described above for nonparametric skew.

Figure 2. Nonparametric skew applied to four altmetric indicators and 22 scientific fields

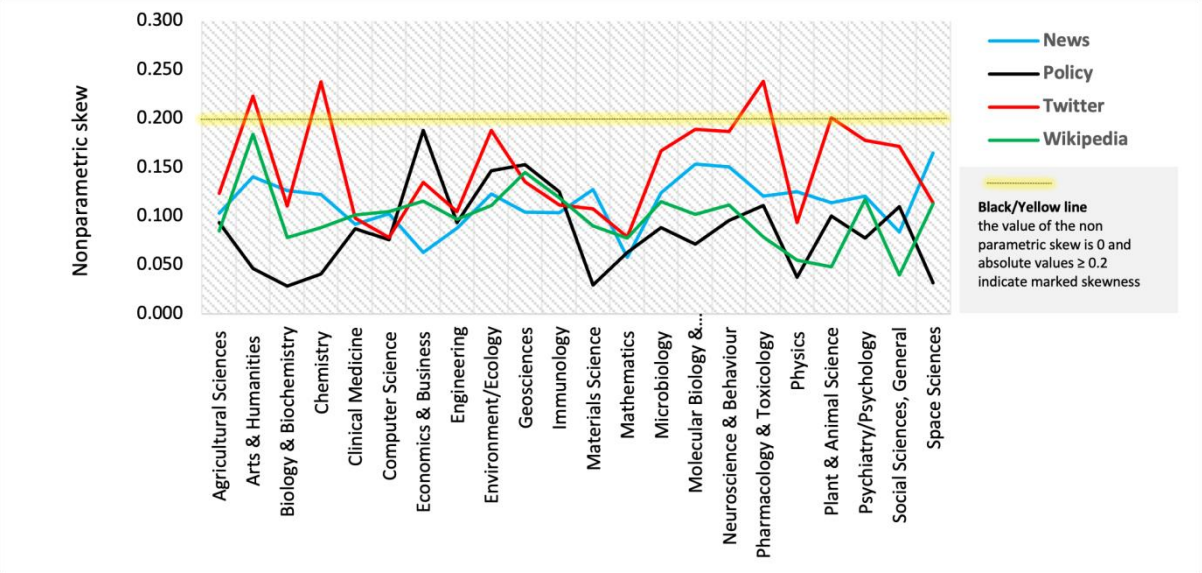


Table 3. Nonparametric skew applied to four altmetric indicators and 22 fields

Scientific Field	News Mentions	Policy Mentions	Twitter Mentions	Wikipedia Mentions
Agricultural Sciences	0.1030	0.0933	0.1238	0.0851
Arts & Humanities	0.1403	0.0468	0.2233	0.1838
Biology & Biochemistry	0.1263	0.0286	0.1105	0.0781
Chemistry	0.1223	0.0412	0.2379	0.0885
Clinical Medicine	0.0918	0.0872	0.0979	0.1012
Computer Science	0.1024	0.0761	0.0785	0.1048
Economics & Business	0.0629	0.1879	0.1350	0.1155
Engineering	0.0881	0.0933	0.1047	0.0967
Environment/Ecology	0.1228	0.1466	0.1878	0.1112
Geosciences	0.1041	0.1527	0.1354	0.1453
Immunology	0.1034	0.1251	0.1117	0.1188
Materials Science	0.1272	0.0297	0.1078	0.0902
Mathematics	0.0581	0.0630	0.0785	0.0776
Microbiology	0.1244	0.0886	0.1669	0.1151
Molecular Biology & Genetics	0.1537	0.0717	0.1891	0.1019
Neuroscience & Behaviour	0.1506	0.0956	0.1867	0.1114
Pharmacology & Toxicology	0.1207	0.1110	0.2386	0.0785
Physics	0.1251	0.0373	0.0937	0.0549
Plant & Animal Science	0.1140	0.1006	0.2008	0.0482
Psychiatry/Psychology	0.1208	0.0779	0.1781	0.1171
Social Sciences, General	0.0841	0.1100	0.1717	0.0399
Space Sciences	0.1648	0.0319	0.1140	0.1130

The case of policy mentions is singular. This altmetric indicator has the scientific fields with the lowest nonparametric skew. The fields with the lowest asymmetry are "Biology & Biochemistry" (0.0286), "Space Sciences" (0.0319) and "Materials Science Policy" (0.0297).

One exception in the case of policy mentions is "Economics & Business" (0.1879). As for Wikipedia, "Arts & Humanities" has the highest value with 0.1838 and "Social Sciences, General" the lowest with 0.0399. Another interesting aspect is that if we look at a single scientific area, it may have a very different pattern for each of the four indicators. For example, "Chemistry" has strong asymmetry and one of the highest values (0.2379) in the Twitter mentions indicator. However, in the policy mentions indicator it has one of the lowest nonparametric skew values (0.0412). Such extreme patterns can be clearly observed in other scientific fields such as "Pharmacology & Toxicology" and "Arts & Humanities". In the case of the Gini index, the results and patterns are very similar to those described above for nonparametric skew.

Figure 4. Nonparametric skew applied to citations (Dimensions) and 22 fields

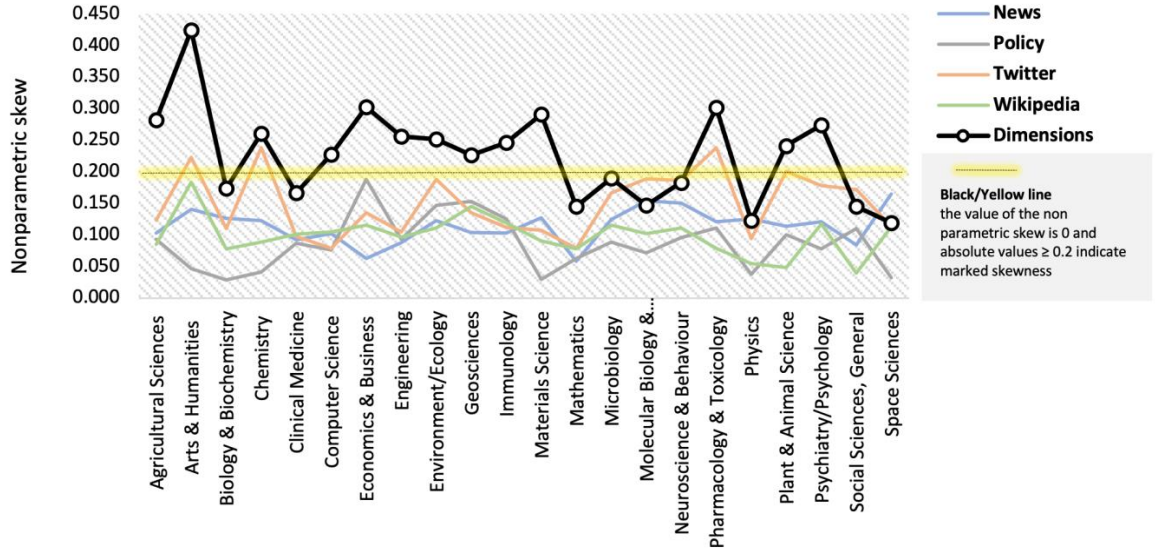


Figure 4 shows the comparison of the Dimensions citations with the four altmetric indicators corresponding to objective 3. The overall nonparametric skew indicator of the citations is 0.164. This is much higher than the four altmetric indicators selected. This pattern is repeated for the 22 scientific fields, since in almost all cases with the exception of "Molecular Biology & Genetics" and "Social Sciences, General" on Twitter, the citations have much higher nonparametric skew values. There are fields where the asymmetry of citations is pronounced or very pronounced, such as "Arts & Humanities" (0.4243), "Economics & Business" (0.3023) and "Materials Sciences" (0.2910), always with significant differences in the altmetrics. Finally, citations have similar nonparametric skew values in line with those of Twitter in fields such as "Chemistry" (0.2606), "Neuroscience & Behaviour" (0.1831) and "Microbiology" (0.1902).

Conclusions

First of all, it should be mentioned that although this study uses a large sample of scientific publications (a total of 237,232), they are all Spanish publications. Therefore, the results may not be extrapolated. Recent studies show that altmetric indicators may have different patterns depending on the country. Countries such as Spain, France and Germany may have different altmetric patterns to Anglo-Saxon countries (Torres-Salinas et al., 2022). However, various findings have been demonstrated that can be extrapolated to all contexts regardless of their specific features. Each altmetric indicator has its own pattern of asymmetry and is not the same in all scientific areas. The values are very different depending on the area and the indicator. Another important aspect pointed out by the study is that, compared to citations, the

distributions of altmetric indicators are always less skewed and less pronounced. It also seems that citations are more similar to Twitter mentions. This paper is useful to provide a general mapping by indicator and area of the phenomenon of asymmetry in the world of altmetrics. It may be of use when establishing the field validity of certain indicators or when using statistical indicators such as averages. It will also help to decide whether it is necessary to introduce standardisation procedures for indicators such as those used by Costas (2017) and Bornmam and Leydesdorff (2018). This work will be continued in the future using the complete Altmetric.com database and introducing a larger number of altmetric indicators.

Bibliography

Abramo, G., D'Angelo, C. A., & Soldatenkova, A. (2017). An investigation on the skewness patterns and fractal nature of research productivity distributions at field and discipline level. *Journal of Informetrics*, *11*(1), 324-335. https://doi.org/10.1016/j.joi.2017.02.001

Bornmann, L., & Leydesdorff, L. (2017). Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *Journal of Informetrics*, *11*(1), 164-175. https://doi.org/10.1016/j.joi.2016.12.001

Costas, R., Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2017). On the quest for currencies of science: Field "exchange rates" for citations and Mendeley readership. *Aslib Journal of Information Management*, *69*(5), 557-575. https://doi.org/10.1108/AJIM-01-2017-0023

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

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317-323.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45), 17268-17272. https://doi.org/10.1073/pnas.0806977105

Schmoch, U. (2020). Mean values of skewed distributions in the bibliometric assessment of research units. *Scientometrics*, *125*(2), 925-935. https://doi.org/10.1007/s11192-020-03476-8

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628-638. 1097-4571(199210)43:9

Tan, F. (2020). Mapping of Subject category (SC) field—ESI discipline category. Retrieved from https://figshare.com/articles/dataset/Mapping_of_Subject_category_SC_field_-_ESI_discipline_category/13269737

Thelwall, M., & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*, *67*(8), 1962-1972. https://doi.org/10.1002/asi.23501

Torres-Salinas, D., Robinson-Garcia, N., Miguel Campanario, J., & Delgado López-Cózar, E. (2014). Coverage, field specialization and the impact of scientific publishers indexed in the

Book Citation Index. *Online Information Review*, *38*(1), 24-42. https://doi.org/10.1108/OIR-10-2012-0169

Torres-Salinas, D., Robinson-Garcia, N., & Arroyo-Machado, W. (2022). Coverage and distribution of altmetric mentions in Spain: a cross-country comparison in 22 research fields. *El Profesional de la Información*, *31*(2). https://doi.org/10.3145/epi.2022.mar.20

van Raan, A. F. J. (2006). Statistical properties of bibliometric indicators: Research group indicator distributions and correlations. *Journal of the American Society for Information Science and Technology*, *57*(3), 408-430. https://doi.org/10.1002/asi.20284

Waltman, L., van Eck, N. J., & van Raan, A. F. J. (2012). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, *63*(1), 72-77. https://doi.org/10.1002/asi.21671

Yu, H., Xu, S., Xiao, T., Hemminger, B. M., & Yang, S. (2017). Global science discussed in local altmetrics: Weibo and its comparison with Twitter. *Journal of Informetrics*, *11*(2), 466-482. https://doi.org/10.1016/j.joi.2017.02.011

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