Publications in Open Access Journals: A Bibliometric Analysis

Patrícia Mascarenhas Dias(✉), Thiago Magela Rodrigues Dias, and Gray Farias Moita

Federal Center for Technological Education of Minas Gerais, Belo Horizonte, Brazil
patriciamdias@gmail.com

Abstract. The results of scientific activities and consequently their dissemination, mainly through scientific articles, make it possible to publish the techniques and indicators used, as well as the results obtained during development, to various interested parties in the investigated themes. Therefore, identifying how the results of scientific research are being published allows us to understand how scientific communication has been used to disseminate the studies carried out. In this context, the publication of articles in open access journals emerges as an interesting mechanism for the dissemination of scientific research, as it facilitates and enables access to them, since there are no barriers, especially financial ones, for access to content. of this type of publication. Therefore, this study aims to analyze, with the help of different indicators, the entire set of publications in open access journals, registered in the Lattes Platform curricula. In view of the extraction and treatment of curricular data, and with the validation with data extracted from the DOAJ, it was possible to verify the impact of the scientific journals in which the articles are being published, as well as the average of citations of the publications over the years and considering the different areas of knowledge.

Keywords: Open access · Qualitative analysis · Lattes platform

1 Introduction

According to several studies, journals – especially those available in electronic format – have been on the rise since the last decade. It can be said that journals, in all areas of knowledge, have the role of being a filter for the recognition of works that have been accepted. For [1], publication in a journal recognized by the area is the most accepted way to register the originality of the work and to confirm that the works were reliable enough to overcome the skepticism of the scientific community.

In view of the above, the Open Access Movement emerged in the early years of the 21st century, aiming to make the full text of publications available, for use without financial, legal or technical barriers that are not inseparable from access to an Internet connection [2].

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
A. L. Pinto and R. Arencibia-Jorge (Eds.): DIONE 2022, LNICST 452, pp. 60–69, 2022.
https://doi.org/10.1007/978-3-031-22324-2_6

In [3] presents figures that justify and are great motivators for publication in open access journals. For the author, open access articles have 336% more citations than articles that are available in print, with a median equal to 158%.

Therefore, access to scientific information has been a major challenge for developing countries such as Brazil. With the crisis of journals, which arose as a result of the high costs of maintaining subscriptions to scientific journals, access to this type of information was quite limited. Although this crisis began in the mid-1980s, there is still no definitive solution. With information and communication technologies, the open files initiative appears, which defines a model of interoperability between libraries and digital repositories, enabling alternatives for scientific communication [4].

It is also possible to notice that in most cases public resources are present in the stages of formation of a researcher. As a student, the researcher is generally supported by public schools and universities or receives public subsidies to assist in their training, such as scholarships to encourage research. As a researcher, he receives public funding to develop his research and support for participation in national and international scientific events to present research results. Finally, once published, the research results depend on the use of public resources to acquire the journals in which these researches are found. As they are scientific journals that bring together several articles in a single volume, it is necessary to acquire the set of articles published in those journals, which does not always reflect the needs of the researcher or funded research institutions [5].

In Brazil, the CAPES Periodicals Portal is one of the main sources of access to data from scientific publications. In it, it is possible to access a collection containing more than 45 thousand titles with full text, 130 reference bases and 12 bases dedicated exclusively to patents, in addition to books, encyclopedias and reference works, technical standards, statistics and audiovisual content [6].

However, the CAPES Periodicals Portal is not characterized as an open access initiative, since the government agency pays for subscriptions to periodicals made available free of charge to member institutions. In 2018 alone, the cost was BRL 402 million [7].

Investment in the CAPES Periodicals Portal has been increasing. In 2014, there was a decrease of approximately 3.3% compared to the previous year. During the years 2017 and 2018, the investment was practically the same, and according to CAPES the forecast for investing in the year 2020 in the portal was approximately 347 million reais. When considering the period between 2004 and 2019, the investment already made by CAPES to maintain the portal reaches approximately 3.1 billion reais.

Therefore, this work uses curriculum data from the Lattes Platform as a source of information for analysis using metrics such as Impact Factor and number of citations in order to better understand the set of articles published in open access journals by Brazilian researchers.

This type of study is characterized as an important mechanism to evaluate the evolution of publications in open access journals by Brazilian researchers with the adoption of qualitative indicators, making it possible to verify whether the incentive policies for the publication of research in this communication format have achieved results satisfactory.

2 Methodology

For the analysis of Brazilian scientific production in open access journals, the set of all individuals with CVs registered on the Lattes Platform was considered. The selection of this data set for the analysis is motivated by the fact that most Brazilian researchers have registered CVs and even researchers at the beginning of their careers are encouraged to register and keep their CVs up to date. Thus, the group of individuals analyzed in this work comprises a large part of the researchers working in Brazil.

In view of the difficulty in collecting and manipulating the set of curricula that make up the Lattes Platform, LattesDataXplorer [8], a framework for extracting and data processing was used (Fig. 1).

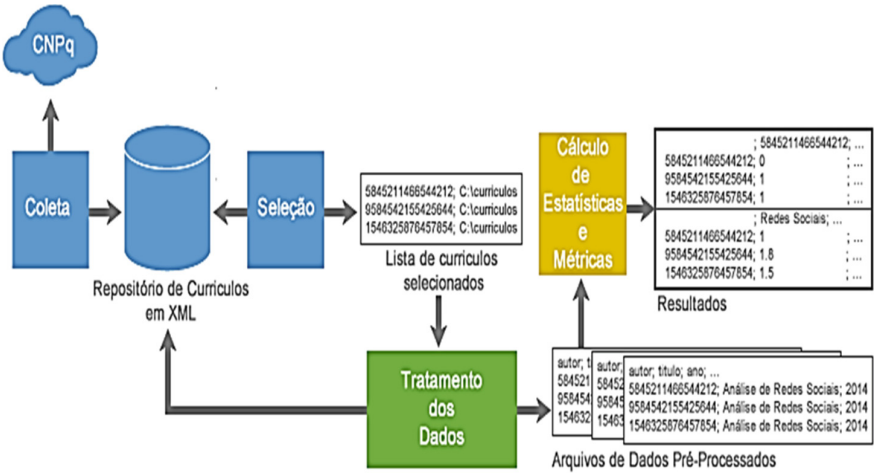

Fig. 1. LattesDataXplorer architecture [8].

A curriculum registered on the Lattes Platform can contain a lot of information capable of helping to understand the evolution of Brazilian science from different perspectives. However, for the purposes of this work, only data from articles published in open access journals, as well as their authors, were taken into account. Given the above, new components for the original framework were developed to meet the specific demand to identify for each article published in a journal (namely, 7,841,860), of each of the individuals (namely, 6,548,210) (data collected in January 2020), whether the journal in which that article had been published was open access (Fig. 2).

After collecting the entire repository of curricula using the framework initially proposed for this purpose, the new components developed were used to identify all records of open access journals from the Directory of Open Access Journals (DOAJ). As of January 2020, the DOAJ indexed 14,170 journals and 4,543,939 articles. DOAJ has been a data source and reference on open access journals for several studies [9–13].

By crossing and validating the data of each article registered in a given curriculum with the list of ISSN and eISSN of the DOAJ, it was possible to identify the publications

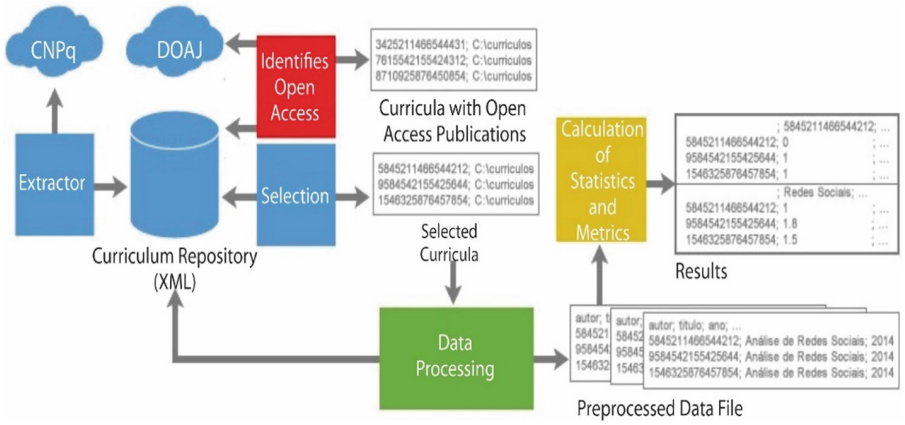

Fig. 2. LattesDataXplorer extended.

considered in this research as being an article in an open access journal. Therefore, all data from the publication, as well as from the analyzed curriculum, were included in an extract of valid data. This entire process occurs for all publications considered from all curricula initially collected.

By analyzing all article records from the collected dataset (7,841,860), 29.92% (2,090,015) was published in open access journals.

As already described, articles in open access journals have been the subject of interest for several researches in various areas of knowledge. Such interest is evident when analyzing the growing volume of publications over the years (Fig. 3).

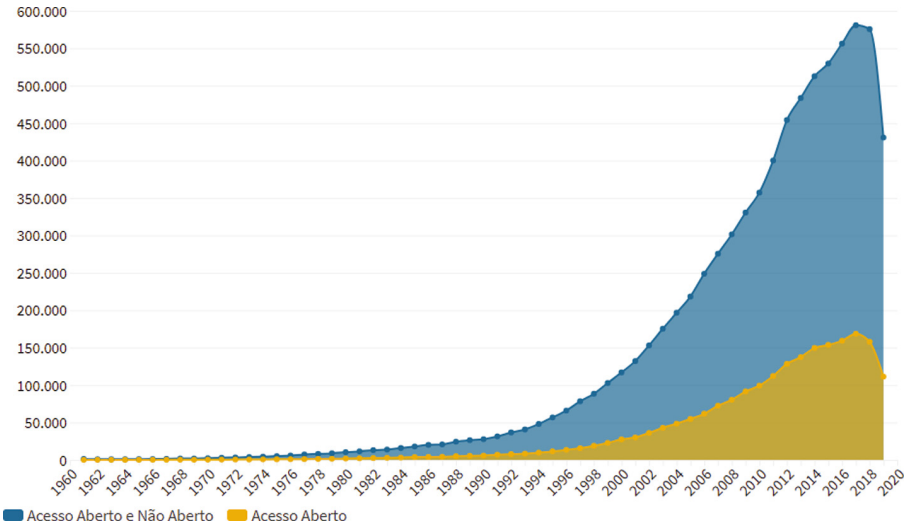

Fig. 3. Temporal comparison of the analyzed publications.

It is possible to observe a constant growth in the quantity of publications in periodicals, where in the year 2000, the percentage of publications in open access periodicals represented approximately 24%. The peak of representation of articles in open access journals occurs in 2017 with a percentage of 29.02%.

3 Analysis and Results

An important piece of information to understand the profile of authors who have already published in open access journals concerns their level of academic training. Such information is important because it makes it possible to understand the academic profile of authors who have published in journals of this format in the country. In a curriculum registered on the Lattes Platform, it is possible to record the entire history of an individual's completed and ongoing academic training, from elementary school to doctorate.

Although the postdoctoral degree is not a level of training, the authors who did it were separated from the group of doctors for statistical purposes and better understanding. It can be seen that most authors have a completed doctorate (47.62%). In second place, there are authors with a master's degree (22.51%), followed by those who completed post-doctoral studies (18.12%), so that only these three groups account for 88.25% of the analyzed group. In addition, it is observed that only 0.14% of the authors did not report completed training levels in their curricula, and that 4.56% only have completed graduation. The absence and low representation of authors at the lowest levels of academic training can be explained by the fact that scientific publications, especially those carried out in journals, are, in general, the result of research carried out in postgraduate studies.

In view of this, it is evident that most of the set of authors identified for having already published works in open access journals is composed of individuals with a high level of academic training. This justifies its significant representation in the total set of publications of articles in journals, even if it is a small percentage of the total set of individuals with CVs registered on the Lattes Platform.

According to [8], there is a tendency for researchers with higher levels of training to be more productive. This is due to the need to publicize their research, which is usually carried out in their postgraduate training process or in collaborations originated from orientations also carried out in postgraduate studies. Considering that the number of publications has been increasing year after year, driven mainly by the ease of dissemination and access to information through the web, and by the high degree of academic training of the group of authors analyzed in this work, several bibliometric metrics can be applied in order to get new results.

In this context, the importance of analysis of articles published in journals is evident, as they are one of the main means of disseminating research results. In addition, articles published in journals are also frequently used as indicators for the evaluation of researchers, research institutions and even graduate programs. Therefore, in the context of this research, which aims to analyze publications in open access journals in Brazil, performing an analysis with qualitative metrics in this means of dissemination is necessary.

The Impact Factor (IF) proposed by Garfield in 1955 is considered very valuable information for qualitative analyses. Currently, the IF is published annually by the Journal

of Citation Report – JCR, as calculated by Thomson Reuters. To calculate the IF of a particular journal in a given year, the number of citations received in that year by articles published in the journal in the two preceding years is taken into account, divided by the number of articles published by the journal in the same period. It is worth mentioning that several researchers use the IF of the journal as a selection criterion for publication of their articles so that their work has greater visibility, hence the importance of the IF to qualify the journal.

Evaluating the IF of the open access journals analyzed in this work, Fig. 4 shows the medians in which the articles were published over the years. As can be seen, the apex of the medians occurred in the year 2000 with a value equal to 0.5340. After the year 2000, for three consecutive years, there was a decrease, and later, there was a growth for two years, after, from 2006 onwards, there were several drops until reaching the median equal to 0 in 2010, which remains until the year 2000. Last year analyzed in this work (2019). For better visualization of the results, Fig. 5 presents the distribution of the medians without the outliers, considering that some journals have IF much above the average.

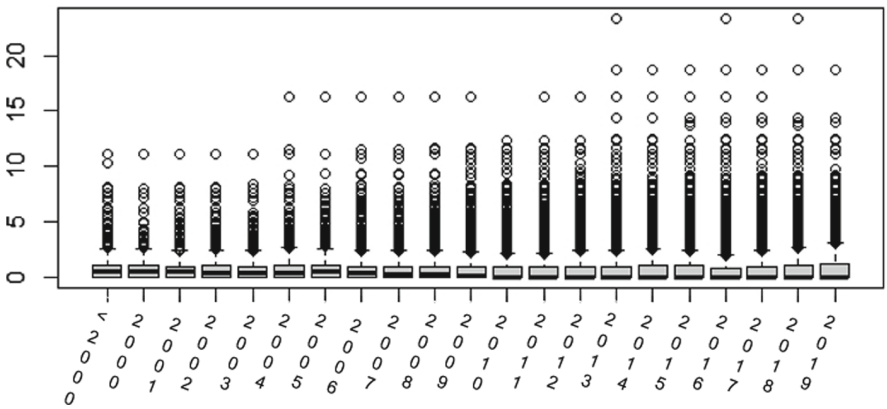

Fig. 4. Distribution of median impact factor over years.

Therefore, it is possible to perceive by the behavior of the medians, that from 2010 onwards, the publication of articles in open access journals analyzed in this work happened mostly in journals with IF equal to 0. It is worth mentioning that, even the publications of articles in open access journals being on the rise, most works are being published in journals with low IF. This fact does not necessarily indicate that publications in good journals are no longer happening, but that a large number of articles are being published in journals with low IF value.

In addition to the IF, which is often used as an indicator to qualify a journal, another indicator widely used to assess the quality of a given work refers to the number of citations that the work receives. To analyze the individual quality of the works considered in this work, the number of citations of each of the works published in open access journals was extracted from the Web of Science (WOS), Scielo and Scopus. Figure 6 shows

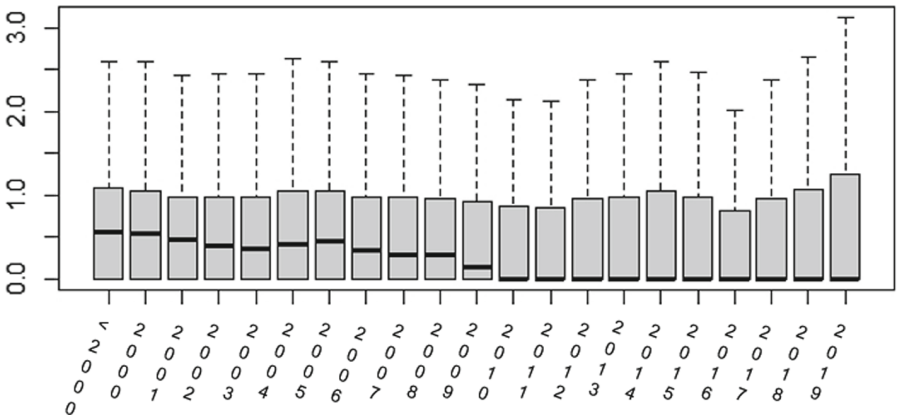

Fig. 5. Distribution of median impact factor excluding outliers.

the average citations of articles published in each of the years in the three directories considered.

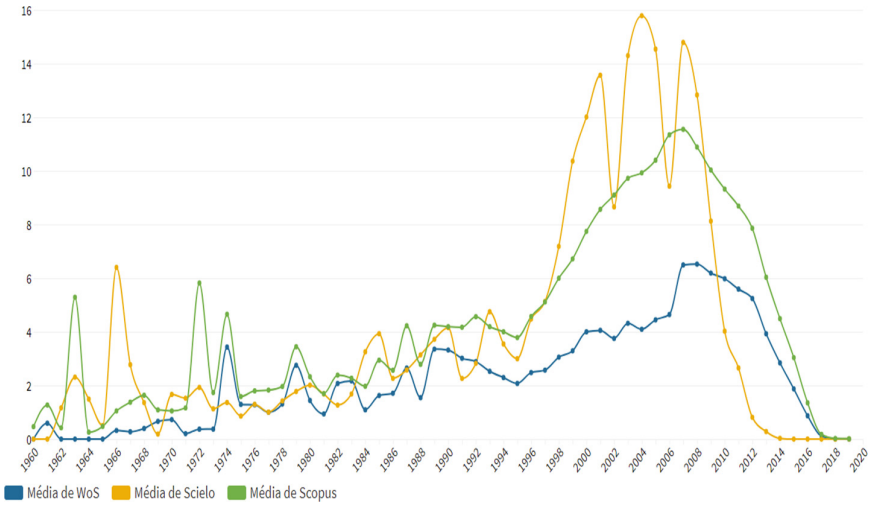

Fig. 6. Average citations of analyzed articles.

As can be seen, the analysis begins in 1960, the year of the first publication of the selected set and, in general, the average of citations of the articles has a variability in the following years in the three directories of citations consulted. A constant increase in the average of citations of articles can be seen from the end of the 1970s onwards. From 1996 onwards, the growth of the average of citations of Scielo becomes much more representative than that of the other directories, with the average of WoS being keeping with more modest growth than the other directories. It is noteworthy that in the articles published in the last years of the analysis, the average of citations tends to

suffer a significant drop, starting to have average values very close to zero from 2017 onwards in the three directories, considering that recently published works tend to not have received many citations.

It is important to highlight that the quality of the scientific production of a given researcher cannot be measured simply by the number of citations that their work has received. This is due to several factors, such as, for example, his career time or the collaboration network in which he is inserted [8]. The number of references per article can be quite different in each of the different areas of knowledge. For example, articles in the Exact Sciences generally have a lower citation density than those related to the Health Sciences, which explains, in part, why the IFs of journals in this area are, on average, much higher compared to those of the Exact Sciences, as in the case of Mathematics [14].

The large discrepancy in the number of citations received by articles published in open access journals in different areas of knowledge makes a more detailed analysis necessary to make it possible to understand this distinction.

4 Conclusions

Currently, several studies have used data from scientific publications as a source for different types of analysis [15–17]. The publication of articles in open access journals has attracted the interest of the entire scientific community, in view of the growth in the number of journals in this format, as well as the recognized relevance that works in this publication medium have acquired. Allied to this, there is the interest of the scientific community in verifying the efforts made to boost the publication of articles in this publication medium, mainly due to the financial costs involved in accessing research that is usually financed with public resources. For these reasons, it is important to understand how publications in this dissemination format are evolving.

When carrying out an analysis that considered traditional quality metrics, articles published in open access journals registered in the Lattes Platform curricula were analyzed, considering their citations in WoS, Scielo and Scopus, as well as the IF of the periodicals. The IF is considered valuable information for qualitative analyses. Therefore, evaluating the IF of open access journals, a great distinction was observed in their values, ranging between 0 and 23.333. Therefore, when performing a temporal analysis of the articles, considering the median of the IF's of all journals, it was identified that the apex occurred in the year 2000, with a median value equal to 0.5340. It is noteworthy that in 2010 the median value was equal to zero, a value that remains until the last year analyzed in this work (2019). A hypothesis for this median value equal to zero, from the year 2010, may be related to the large number of articles, registered in the curricula from this year, in which, in most publications, the journal has null IF, or that is, with a value of zero. Of the total set of journals analyzed, approximately 82% have an IF equal to zero.

Regarding the citations received by the analyzed articles, an increase in the average of citations of the works published from the end of the 1970s can be seen, and this growth intensified in the late 1990s. Subsequently, a significant drop in the average of citations from 2008, in all considered directories. It is also important to note that as of 2017, in

all directories, the average values of citations are close to zero, considering that recently published works may not have received significant amounts of citations.

The analysis of the average of citations is directly influenced by highly cited articles, which ends up raising such averages, especially in some areas, since even with many articles without citations, these highly cited works end up influencing the results.

In addition, it was possible to observe that articles published in journals with higher IF value tend to receive a greater number of citations, as a result of the credibility and visibility that such journals provide. In general, when published in journals with significant IF value, the work tends to have greater repercussion, to the detriment of most publications that are published in journals with null IF.

References

1. Rodrigues, R.S., Oliveira, A.B.: Periódicos científicos na america latina: títulos em acesso aberto indexados no ISI e SCOPUS. Perspectiv. Ciênc. Inf. **17**(4), 76–99 (2012)
2. BOAI (Budapest Open Access Initiative): Dez anos da Iniciativa de Budapeste em Acesso Aberto: a abertura como caminho a seguir. http://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation (2012). Accessed: Jan 2022
3. Lawrence, S.: Free online availability substantially increases a paper's impact. Nature **411**(6837), 521 (2001). https://doi.org/10.1038/35079151
4. Kuramoto, H.: Informação científica: proposta de um novo modelo para o Brasil. Ciência da Informação **35**(2), 91–102 (2006)
5. Freire, J.D.: CNPq e o acesso aberto à informação científica. 2011. 275 f. Thesis – Curso de Programa de Pós-graduação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília (2011)
6. CAPES: O Portal de Periódicos da Capes. www.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pcontent&alias=missao-objetivos&Itemid=109. Accessed 2022
7. CAPES: Orçamento – Evolução em reais. https://www.capes.gov.br/orcamento-evolucao-em-reais#A. Accessed 2020
8. Dias, T.M.R.: Um Estudo Sobre a Produção Científica Brasileira a partir de dados da Plataforma Lattes. 2016. 181 f. Thesis – Curso de Programa de Pós-graduação, Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte (2016)
9. Lone, F., Rather, R., Shah, G.J.: Indian contribution to open access literature: a case study of DOAJ & OpenDOAR. Chin. Librarianship: an Int. Electron. J. **26**, 1–10 (2008)
10. Swan, A.: Why open access for Brazil? | Por que acesso livre no Brazil. Liinc em Revista **1**(1), 158–171 (2008). https://doi.org/10.18617/liinc.v4i2.279
11. Björk, B.-C., Solomon, D.: Article processing charges in OA journals: relationship between price and quality. Scientometrics **103**(2), 373–385 (2015). https://doi.org/10.1007/s11192-015-1556-z
12. dos Santos Costa, E.H.: Periódicos de acesso aberto: um canal de disseminação dos pesquisadores bolsistas PQ1A do CNPq? 2018. 127 f. Thesis – Curso de Programa de Pós-graduação em Ciência da Informação, Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2018)
13. Sarvo, D.D.O.: Avaliação da Plataforma Lattes como fonte de coleta de metadados para povoamento automatizado em repositórios institucionais. 2018. 96 f. Thesis (Master's degree) - Curso de Programa de Pós-graduação em Ciência, Tecnologia e Sociedade, Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos (2018)

14. Garfield, E.: The use of JCR and JPI in measuring short and long term journal impact. In: Council of Scientific Editors Annual Meeting (2000)
15. Jorge, R.A., González, M.J.P.: Recommendations on the use of Scopus for the study of Information Sciences in Latin America. Iberoamerican J. Sci. Meas. Commun. **1**(1), 8 (2020). https://doi.org/10.47909/ijsmc.07
16. Mascarenhas, H., Dias, T.M.R., Dias, P.: Academic mobility of doctoral students in Brazil: An analysis based on Lattes Platform. Iberoamerican J. Sci. Meas. Commun. **1**(3), 1–15 (2021). https://doi.org/10.47909/ijsmc.53
17. Hernandez-Cruz, N.: Mapping the thematic evolution in Communication over the first two decades from the 21st century: a longitudinal approach. Iberoamerican J. Sci. Meas. Commun. **1**(3), 1–10 (2021). https://doi.org/10.47909/ijsmc.88