Revista ARACÊ
ISSN: 2358-2472

# BIBLIOMETRIC ANALYSIS: LITERATURE REVIEW AND PROPOSAL OF A METHODOLOGICAL FRAMEWORK IN 12 STEPS

https://doi.org/10.56238/arev6n4-146

**Submitted on:** 11/10/2024 **Publication date:** 12/10/2024

**Marcelo Salvador Celestino[1], Regina Celia Baptista Belluzzo[2], João Pedro Albino[3] and Vânia Cristina Pires Nogueira Valente[4].**

## ABSTRACT

Bibliometric analysis is a methodology for retrospective, quantitative, statistical, and descriptive study of the scientific literature, which allows the identification of the most prolific and influential elements in a given field, the interconnections between authors and the literature itself, in addition to identifying emerging research themes. The present study is divided into two parts. First, a narrative review of the literature was presented on the foundations, main techniques and challenges that permeate the conduction of a bibliometric analysis, with the objective of elucidating the theme, serving as a starting point for research of this nature. To illustrate the examples, a bibliometric analysis was simulated with the theme Ionizing Radiation Dose Reduction and Protocol Optimization in Computed Tomography (CT), on an experimental basis, using the *bibliometrix package* in the RStudio software. Then, a methodological framework with 12 steps was proposed, which have been used empirically in the conduction of bibliometric analyses by the present authors. The addition of the simulation concept and a *feedback loop* allowed us to refine the methodology and obtain more reliable results. This study contributes to the literature by proposing a practice-oriented methodological empirical roadmap, which promotes transparency, reproducibility and methodological rigor in bibliometric analyses, promoting open and collaborative science.

**Keywords:** Bibliometric analysis, Infometrics, Media and Technology, Research methodology, Simulation.

[1] Doctorate student and Master in Media and Technology (UNESP)
PPGMiT-DO/FAAC, São Paulo State University "Júlio de Mesquita Filho"
E-mail: marcelo.salvador@unesp.br
ORCID: https://orcid.org/0000-0002-5551-8793
[2] Dr. in Communication Sciences (USP)
Volunteer Professor, PPGMiT/FAAC,
São Paulo State University "Júlio de Mesquita Filho"
Email: rbelluzzo@gmail.com
ORCID: https://orcid.org/0000-0001-9514-2930
[3] Dr. in Business Administration, Quantitative Methods and Informatics (USP)
Volunteer Associate Professor, PPGMiT/FAAC,
São Paulo State University "Júlio de Mesquita Filho"
E-mail: jp.albino@unesp.br
ORCID: https://orcid.org/0000-0001-5965-1869
[4] Associate Professor in Graphic Representation (UNESP), Dr. in Civil Engineering (USP)
Associate Professor and Coordinator of PPGMiT-DO/FAAC,
São Paulo State University "Júlio de Mesquita Filho"
E-mail: vania.valente@unesp.br
ORCID: https://orcid.org/0000-0001-6563-2402

Revista ARACÊ
ISSN: 2358-2472

## INTRODUCTION

Bibliometric analysis is a methodology used by informetria to quantify the scientific production of a given field, through statistical techniques (Ball, 2017). From the perspective of Hjørland (2004), bibliometric studies are tools that, through the mapping of knowledge domains and discursive communities, allow us to analyze how knowledge is structured and organized in different areas. These tools, inserted in the context of domain analysis, have as their main contribution the exploration of the flow of knowledge among peers in the academic-scientific universe, as well as its production and dissemination process in society.

Bibliometric analysis has been used in research in various areas of knowledge, focusing on the most varied themes, such as governance and sustainability (N'ze; Tenkoul, 2024), Artificial Intelligence (AI) in Sign Language Recognition (Zhang *et al.*, 2024) and global cancer research (Karger; Kureljusic, 2023). This versatility in applications shows that it is a strategy of broad academic and business interest.

There are, in the literature, several guides and manuals that describe the particularities in conducting a bibliometric analysis (Aria; Cuccurullo, 2017; Donthu *et al.*, 2021; Uribe; Contreras; Guerrero, 2023; Öztürk; Kocaman; Kanbach, 2024; Passas, 2024), which can represent a challenge for researchers or students seeking their first contact with this methodology. The objectives of this article were: **(i)** to present a narrative review of the literature, exploring the concepts, and main challenges of a bibliometric analysis, with the presentation of illustrated examples of selected techniques; and **(ii)** to propose a methodological framework based on 12 steps, organized in three modules, to serve as a starting point for students and researchers who wish to understand the main fundamentals of this approach in an objective and practical way.

## METHODOLOGICAL PROCEDURES

The study adopts a methodological mix composed of a narrative review of the literature (Green; Johnson; Adams, 2006; Ferrari, 2015) and an exploratory theoretical-conceptual approach, based on the interactive model of Miles and Huberman (1994) (Period of data collection, data reduction, data display, elaboration and verification of conclusions). The integration of both methodologies, added to the experiences of the present authors, supports the proposition of a *structured framework* for bibliometric analysis. The material used in the narrative review was selected for convenience, using

Revista ARACÊ
ISSN: 2358-2472

various sources, such as articles and manuals, without a defined time frame, but covering classic literature to address bibliometric laws.

## ILLUSTRATED EXAMPLES

To illustrate the selected concepts, a simulation of bibliometric analysis was conducted with the theme Ionizing Radiation Dose Reduction and Protocol Optimization in Computed Tomography (CT) (a topic of interest to one of the authors). A search was carried out in the Scopus database, including only articles published in English, from 1998 to 2022. The search *query* used in the Scopus database in *all fields* is presented in **chart 1**.

**Table 1 –** *Search query* used in the Scopus database

| |
|---|
| (optimization OR "Scan Optimization" OR "Dose, Reduction") AND ("Tomography, Spiral Computed" OR "Computed Tomography, Spiral" OR "Computer-Assisted Tomography, Spiral" OR "Computer Assisted Tomography, Spiral" OR "Spiral Computer-Assisted Tomography" OR "Tomography, Spiral Computer-Assisted" OR "Computerized Tomography, Spiral" OR "Spiral Computerized Tomography" OR "CT Scan, Spiral" OR "Scan, Spiral CT" OR "Scans, Spiral CT" OR "Spiral CT Scan" OR "Spiral CT Scans" OR "Helical CT" OR "CT, Helical" OR "CTs, Helical" OR "Helical CTs" OR "Spiral Computed Tomography" OR "Spiral CT" OR "CT, Spiral" OR "CTs, Spiral" OR "Spiral CTs" OR "Tomography, Helical Computed" OR "CAT Scan, Spiral" OR "CAT Scans, Spiral" OR "Scans, Spiral CAT" OR "Spiral CAT Scans" OR "Helical Computed Tomography" OR "Computed Tomography, Helical" OR "Multislice, Computed Tomography" OR "CT, Angiography") AND NOT ("Magnetic Resonance Imaging" OR MRI OR "Positron Emission Tomography" OR PET OR "Radiation Theraphy" OR "radiotherapy" OR "tomotherapy" OR "Ultrasound" OR "Mammography" OR "Fluoroscopy" OR "Radiography" OR "Industrial" OR "Orthodontic" OR "Orthodontics") |

**Note:** MeSH (*Medical Subject Headings*) terms such as *Tomography, Spiral Computed*, and common terms such as *Scan Optimization* were used to enlarge the sample. **Source:** Prepared by the authors (2024).

For data analysis, the free *bibliometrix package* (Aria; Cuccurullo, 2017) of the R programming language, in the RStudio environment (Posit, 2024). This *software*, considered one of the most complete for scientific mapping (Terra *et al*., 2022), organizes workflows for descriptive and/or visual bibliometric analysis, with integration with statistical methods.

## BASES OF THE *PROPOSED* FRAMEWORK

For the elaboration of the methodological framework, guides and bibliometric analyses conducted in practice were consulted, in order to understand the different ways in which the authors apply this methodology (Pizzani; Da Silva; Hossne, 2010; Vasconcelos, 2014; Aria; Cuccurullo, 2017; Donthu *et al*., 2021; Karger; Kureljusic, 2023; Öztürk; Kocaman; Kanbach, 2024). Unlike the studies used as a reference, a simulation stage with *a feedback loop was added*, which corresponds to the performance of tests from the *search query* to the application of several bibliometric techniques, such as the co-

Revista ARACÊ
ISSN: 2358-2472

occurrence of keywords, in order to explore the integrity of the metadata, the adherence to the search problem and, as suggested by Donthu *et al*. (2021), select the techniques to be applied in advance. One of the benefits of simulation is the possibility of performing tests without additional costs or losses, allowing the early identification of failures and refinement of the original process (Celestino; Valente, 2021). In the present study, this critical approach enabled a more conscious and contextualized analysis of bibliometric data, contributing to a deeper and more reflective understanding of how to refine the methodological process and present more reliable and consistent results.

## LITERATURE REVIEW

### WHAT IS A BIBLIOMETRIC ANALYSIS?

The origin of the term bibliometrics is controversial in the literature, involving Paul Otlet and Alan Pritchard. For Momesso and Noronha (2017), it is more likely that Paul Otlet is considered the author of the term, created in the book *Traité de Documentation* (Otlet, 1934),[5] from a broad and complex perspective of documentation. However, the term was formalized by Alan Pritchard (1969), [6]who described bibliometrics as the application of statistical methods to the analysis of publications. For Alvarenga (1998), bibliometrics offers *insights* into how knowledge is constructed and legitimized, allowing a more critical understanding of contemporary scientific practices.

In this context, bibliometric analysis is understood as a quantitative methodology that examines scientific production retrospectively, using metrics that evaluate the performance and influence of literature. In addition, techniques are used that allow the identification of patterns, collaboration networks, and the exploration of the interrelationships that are formed between the elements involved in scientific production (Donthu *et al*., 2021; Passas, 2024). By quantifying aspects of scholarly communication, bibliometric analysis provides an overview of the dissemination and influence of the knowledge produced. The data can guide strategic decisions on investments in research, education, and public health policies, considering the relationships formed in integrated research networks (Ninkov; Frank; Maggio, 2022; Byl *et al*. 2024).

[5] OTLET, P. *Traité de documentation:* le livre sur le livre: théorie et pratique. Bruxelas: Mundaneum, 1934. Disponível em: http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf. Acesso em: 11 abr. 2016.
[6] PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation*, Leeds, v. 25, n. 4, p. 348-349, 1969.

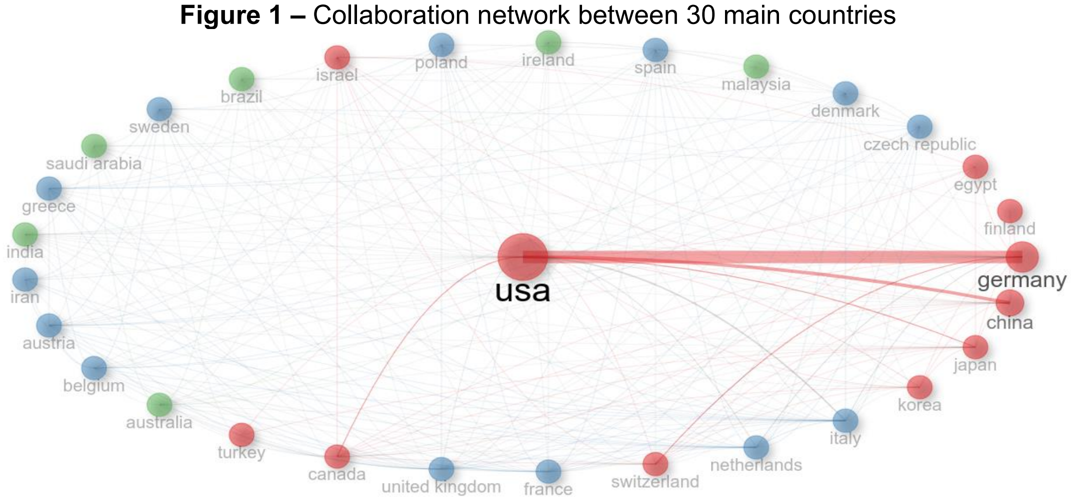
Revista ARACÊ
ISSN: 2358-2472

Öztürk, Kocaman and Kanbach (2024) highlight the common confusion between the terms *research* and bibliometric analysis. They warn of the variation in the approach and quality of studies that use bibliometric analysis, many of which are incorrectly considered as bibliometric research, when in fact they are narrative reviews of the literature. In addition, they observe that, in some cases, bibliometric analysis techniques are applied superficially, without an in-depth evaluation of the results or discussion of their relevance to the specific field of research.

## THE CONCEPT OF NODE

In a network, the node represents the fundamental unit that connects to others, forming a complex structure. The nodes that form in a network can represent any entity, from people in social networks, authors, institutions, to genes in biological networks. Imagine a network as a web: the nodes are the meeting points of the wires, and the lines that connect them represent the relationships between them. By studying the properties of these nodes and their interactions, it is possible to understand the structure and functioning of various networks, such as social and technological networks. Concepts such as *degree* (number of connections), *centrality* (importance in the network), and *clusters* (groups of densely connected nodes) are essential to analyze these complex structures (Newman, 2010). Figure **1** illustrates a network of collaboration formed between countries in the scientific production of simulated bibliometric analysis.

**Figure 1 –** Collaboration network between 30 main countries




**Note:** Collaborative network between 30 countries, formed in research on Ionizing Radiation Dose Reduction and Optimization of CT Protocols. Each circle (node) represents a country participating in the survey, while the lines that connect them indicate the collaborations between them. Clusters were identified according to the colors blue, green, and red, the latter being the most prominent, evidencing a strong collaboration between the United States and Germany. **Source:** Research data analyzed with *bibliometrix* (Aria; Cuccurullo, 2017).

Revista ARACÊ
ISSN: 2358-2472

To analyze visual networks and their communities, algorithms such as *Walktrap* and *Louvain* are used. *Walktrap* simulates short connections between nodes to identify communities, creating hierarchies across network divisions. On the other hand, *Louvain*prioritizes modularity, iteratively regrouping nodes until it achieves the best division (Smith *et al*., 2020; Mukerjee, 2021). While *Walktrap* highlights small communities, *Louvain* is more efficient in large chains, focusing on optimal overall division.

## METRICS AND INDICATORS

Geisler (2000, p. 48) defines the term metrics as a system of measurements that includes the item object of the measurement, the unit of measurement and the value of the unit. In addition, it classifies metrics as objective or subjective. Peer review is cited by this author as a subjective metric, while patent count is considered an objective metric. Also according to this author's definition, metrics can take on various formats, such as a single measure, a ratio (between two measures), an index, or an integrated measure that combines several metrics, including those with different, objective, and subjective attributes.

Metrics are essential for assessing the impact and relevance of published research. However, the evaluation of research should not be reductionist; the analysis must consider multiple metrics (or indicators), supported by expert opinion, in order to obtain *critical insights* and perspectives from academic production (Elsevier, 2009).

While the concept of metric refers to quantitative measures used to evaluate and analyze scientific production, indicators are considered specific measures that represent and quantify more particular characteristics within bibliometrics (Glänzel, 2003). Chart **2** presents a synthesis of examples of metrics evaluated in a bibliometric analysis.

Revista ARACÊ
ISSN: 2358-2472

**Chart 2 –** Examples of metrics

| Metric | Description | Ref. |
|---|---|---|
| Thoughtful Quotes | It gives more weight to citations from influential articles/authors, similar to the *PageRank algorithm*. | [4] |
| Collaboration and co-authorship | Analysis of collaboration between researchers and institutions, which can influence the visibility and impact of publications. | [5] |
| Impact factor | It measures the frequency with which articles in a journal are cited, reflecting their quality and relevance. | [4] |
| *g-index* (índice g) | Index that amplifies the citations of the most cited articles to highlight distinctions between researchers. | [3] |
| *h-index* (índice h) | Metric that provides a balanced view of the author's productivity and impact. E.g., an *h-index* of 10 indicates 10 publications cited at least 10 times. | [2] |
| *m-index* (índice m) | A complementary metric to the *h-index*, which characterizes the influence of the researcher over time. | [2] |
| Total Citations | Counting of how many times the works of an author or institution have been cited by other researchers. | [1] |
| Total publications | Total number of scholarly articles published by an author, group, or institution. | [1] |

**Note:** [1] = Donthu *et al.* (2021); [2] = Hirsch (2005); [3] = Egghe (2006); [4] = Van Noorden (2010); [5] = Elsevier (2009). **Source:** Elaborado pelos autores (2024).

The number of articles published in journals is considered an important factor, since the most productive journals on a given topic can be considered the most relevant within the investigated context, based on Bradford's Law (Alabi, 1979). However, Anthony van Raan, quoted by Van Noorden (2010, p. 864-865), states that "if there is one thing that every bibliometrician agrees on, it is that one should never use the journal's impact factor to evaluate the research performance of an article or an individual — this is a mortal sin."

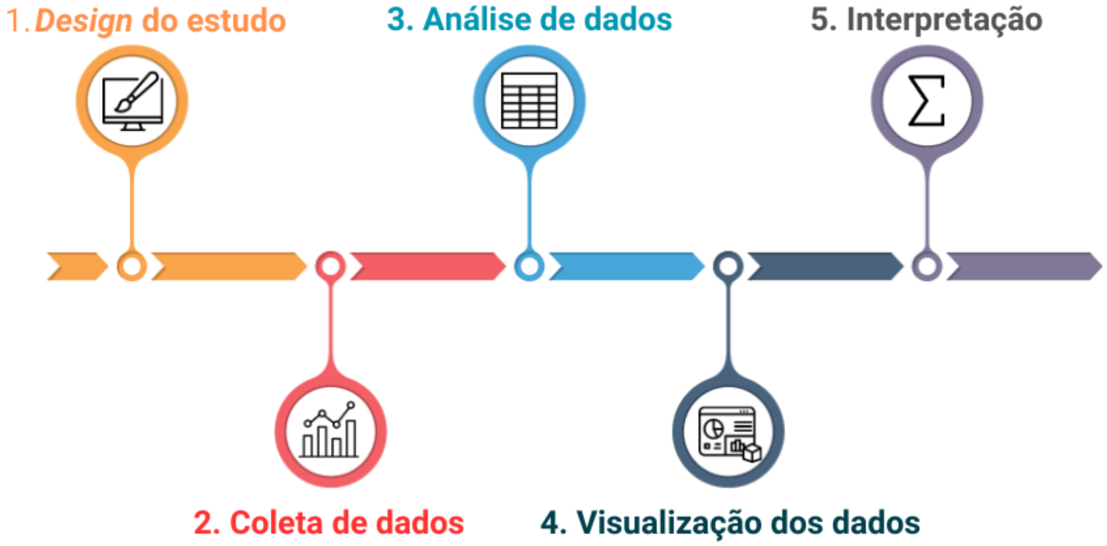
## STEPS IN CONDUCTING A BIBLIOMETRIC ANALYSIS

In general, bibliometric analysis is divided into: **(i)** descriptive analysis; **(ii)** performance analysis; and **(iii)** scientific mapping (Donthu *et al.*, 2021). Performance analysis aims to classify according to the quantification of metrics, while scientific mapping focuses on visualizing the relationships and interconnections that are formed between production and interpretation, and should follow a path capable of associating with gaps in the literature, producing considerations relevant to the area (Öztürk; Kocaman; Kanbach, 2024).

In the book *Introducción a la bibliometría práctica* (Introduction to practical bibliometrics) (Uribe; Contreras; Guerrero, 2023), a structured process for bibliometric analysis is proposed, starting from the definition of the research objective to the presentation of the results. Aria and Cucurullo (2017) present a *workflow* based on Zupic and Čater (2015), which consists of five main steps, as can be seen in **figure 2**.

Revista ARACÊ
ISSN: 2358-2472

**Figure 2 –** *Workflow* for scientific mapping




**Note:** *Design* of the study refers to the methodological framework used in the research. **Translated terms:** 1. Study design, 2. Data collection, 3. Data analysis, 4. Data visualization, 5. Interpretation. **Source:** Prepared by the authors (2024), based on Aria and Cuccurullo (2017) and Zupic and Čater (2015).

Each of the stages can comprise sub-stages, as in step 3 (Data analysis), in which the choice of *software*, data processing or the definition of subgroups to be analyzed are carried out. Aria and Cuccurullo (2017) consider scientific mapping holistically, as a continuous process of bibliometric analysis, used to understand the evolution of knowledge. For the present article, the perspective of Donthu *et al*. (2021), in which scientific mapping is one of the evaluation stages present in bibliometric analysis, which allows for a sequential and more structured presentation of the results.

## PRESENTATION OF RESULTS

One way to start the presentation of the results is through the general data of the sample, which includes the total number of documents, authors, collaboration data, among others (Volpe *et al*. 2023), followed by the analysis of the authors' performance, with the most relevant journals and a presentation of the historical context, mentioning the first published work in the sample (Montazeri *et al.*, 2023). Sequentially, the techniques of scientific mapping are presented, as previously mentioned. Table **1** summarizes the data and compares *datasets* 1 (3331 articles) and 2 (2252 articles) used in the example simulation.

Revista ARACÊ
ISSN: 2358-2472

**Table 1 –** Comparison of general data of the samples

| Key Data Information | *Dataset 1* | *Dataset 2* |
|---|---|---|
| *Timespan* | 1998:2022 | 1998:2022 |
| *Sources* (*Journals, Books, etc.*) | 761 | 496 |
| *Documents* | 3331 | 2252 |
| *Annual growth rate %* | 11,04 | 24,68 |
| *Document average age* | 8,9 | 8,23 |
| *Average citations per doc* | 27,67 | 25,55 |
| *References* | 80633 | 55909 |
| *Keywords plus* | 12342 | 10189 |
| *Author's keywords* | 5272 | 4566 |
| *Authors* | 13122 | 9616 |
| *Authors of single-authored docs* | 73 | 29 |
| *Single-authored docs* | 79 | 34 |
| *Co-authors per doc* (Co-authors per document) | 6,54 | 6,77 |
| *International co-authorships %* | 20,23 | 21,05 |

**Source:** Research data analyzed with *bibliometrix* (Aria; Cuccurullo, 2017).

The overview presented in **Table 1** provides an initial view of the general characteristics of each sample, allowing us to begin the report of the findings with a descriptive analysis of the study (Aria; Cuccurullo, 2017). A brief comparison of the data reveals some discrete variations, such as the *average age of the documents*, but a more than double variation in the *annual growth rate*.

The description of the results will depend on the objective of the study and the information that is intended to be evidenced. For the sake of structure and textual organization, it is suggested that the authors' inferences and considerations about these results be presented in the discussion section, following the same order in which they appeared in the results (Montazeri *et al*., 2023; European Radiology, 2024).

## PERFORMANCE ANALYSIS AND SCIENTIFIC MAPPING TECHNIQUES

**Chart 3** presents two categories and the main techniques used in a bibliometric analysis (Donthu *et al.*, 2021), enabling the researcher to carry out a comprehensive and contextualized analysis of the *corpus* of literature.

**Chart 3 –** Examples of techniques used in bibliometric analysis.

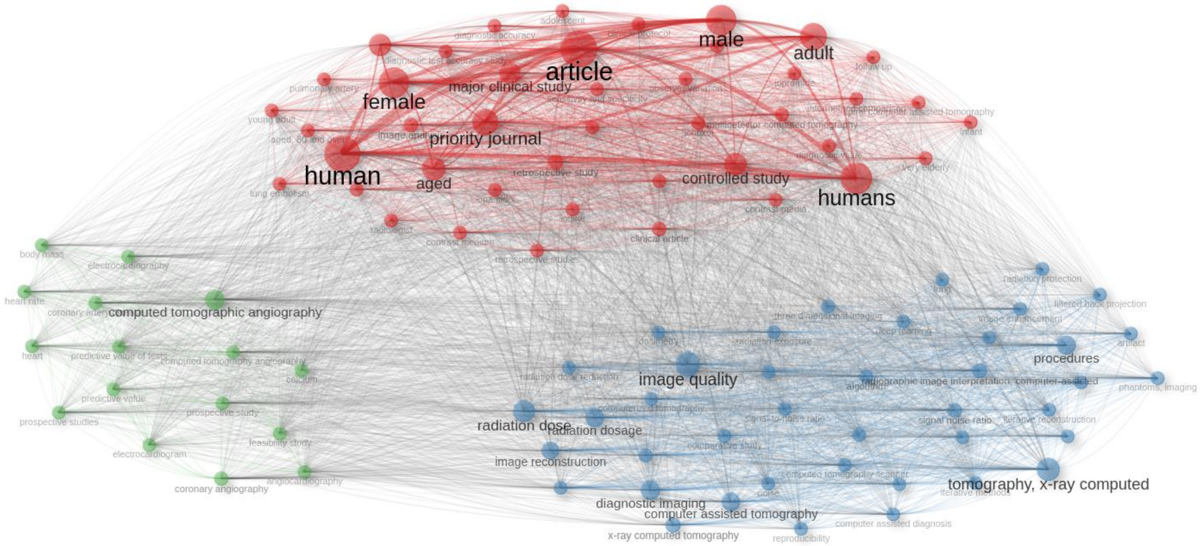
| Category | Techniques | Description | Ref. |
|---|---|---|---|
| **Performance Analysis** | Analysis of production metrics | Evaluation of contributions from authors, institutions, countries, and journals, including publications (TP), and impact metrics (TC, TCm, *h-index*, etc.), as illustrated in **Table 2**. | [1] |
| **Scientific Mapping** | Citation Analysis | Evaluates the influence of publications based on the number of citations they receive. | [2], [3] |
| | Co-Citation Analysis | Measures the similarity between documents, authors, or journals based on how often they are cited together. | [1], [2], [3] |
| | Bibliographic Coupling | It examines the relationship between documents that cite the same references. | [1], [3] |
| | Word co-occurrence analysis | Identifies connections between concepts that occur together in titles, keywords, or abstracts. | [1], [2] |
| | Co-authoring Analysis | It evaluates collaboration between authors based on co-authorship in publications. | [2] |
| | Network Mapping | Visualizes relationships between documents, authors, and institutions using *mapping software*. | [2] |
| | Trend Analysis | It examines the evolution of topics and themes over time to identify emerging areas. | [2] |

**Note:** [1] = Donthu *et al*. (2021); [2] = Aria and Cuccurullo (2017); [3] = Zupic and Čater (2015). Ref. = References; TC = Total citations; TCm = Average citations per publication; TP = Total publications. **Source:** Prepared by the authors (2024).

**Figure 3** presents a co-occurrence network based on 100 general keywords, built with the *Louvain* algorithm. In the image, three *clusters* are identified formed around central topics, which establish connections between them, creating a broad network, in which certain terms often appear together in a set of publications, suggesting a semantic or thematic relationship between them.

Revista ARACÊ
ISSN: 2358-2472

**Figure 3 –** Co-occurrence network of 100 keywords




**Note:** The *red cluster* addresses demographic data, the blue cluster deals with image quality and radiation dose reduction in CT, and the green cluster deals with applications. The size of the circle indicates the frequency of the word, and the thickness of the lines, the strength of the association between the terms.
**Source:** Research data analyzed with *bibliometrix* (Aria; Cuccurullo, 2017).

In addition to the techniques described in **chart 3**, the so-called bibliometric laws can be applied, which, according to Tague-Sutcliffe (1992), established the theoretical bases of information technology, a field that quantifies and analyzes information-related phenomena. These laws are highlighted below.

## BIBLIOMETRIC LAWS

### Lotka's law

Published in 1926, Lotka's Law reveals that the productivity of authors in a specific area follows an unequal distribution. Under this law, a small number of authors are responsible for the majority of publications, while the majority contribute a smaller number of papers. Lotka observed in his analysis that about 60% of the authors had only one publication (Lotka, 1926). In a literature review on information sharing on social media, Abbas *et al*. (2022) evaluated 825 documents published between 2009 and 2020, with a total of 2251 authors. The distribution found also revealed a discrepancy in relation to Lotka's Law, with more than 75% of the authors contributing with only one article. This deviation may suggest that it is an area with high fragmentation or specific research practices. In addition, it is necessary to consider that this law has a general context and may vary over time, since the productivity of authors is not static.

Revista ARACÊ
ISSN: 2358-2472

**Bradford's Law**

Described by Samuel C. Bradford in 1934, Bradford's Law states that most citations in a field of study come from a relatively small number of major journals. Bradford observed that, when ordering journals by number of citations received, the most cited journals (Zone 1 or *Core* Source) accounted for a large proportion of total citations. This proportion gradually decreased as one moved towards the least cited journals (Zone 2 and Zone 3) (Alabi, 1979).

In a bibliometric analysis on innovation models, Guimarães, Moreira, and Bezerra (2021) identified that the sample of 919 journals adhered to Bradford's Law, presenting a coherent distribution according to this law (Zone 1 = 8.27% journals, *n* = 33.16% articles (1/3 of the articles in the sample; Zone 2 = 36.56% periodicals, *n* = 33.88% articles; and Zone 3 = 55.17% periodicals, *n* = 32.96% articles). The distribution of articles is balanced, with a small emphasis on Zone 2. Journals in Zone 1, despite having the smallest share of the sample, are responsible for practically one third of the production, while Zone 3, which contains most of the journals, produced a slightly smaller number of articles. This means that, although Zone 3 comprises most of the journals, Zone 1 is proportionally more efficient in terms of scientific production, concentrating a significant production in a small number of journals.

In the bibliometric analysis of Navarro-Ballester *et al*. (2023), on Covid-19 publications in radiology journals between 2020 and 2021, *European Radiology* appeared as the most productive journal (*n* = 167 articles) and the second in number of articles cited (*n* = 32, 4.95%) in a thematic *cluster*. In the study by Volpe *et al*. (2023), *European Radiology* appeared in second place in productivity (*n* = 255, 4.95%). In the bibliometric analysis on AI for cancer detection by Karger and Kureljusic (2023), this journal appeared as the eighth most productive (*n* = 69). This demonstrates that the productivity or relevance of a journal varies according to the theme and scope investigated.

**Zipf's Law**

Zipf's Law, published in 1949, describes an inverse relationship between the frequency of occurrence of words and their position in the *frequency ranking*. As described in the original excerpt, the frequencies "decrease according to the following simple harmonic series: 1, 1/2, 1/3, ... 1/n, since each frequency, *f*, when multiplied by its *rank*, *r*, will result in a constant" (Zipf, 1949, p.35). This means that in a linguistic *corpus*, words of

Revista ARACÊ
ISSN: 2358-2472

higher frequency appear more frequently, while those of lower frequency follow this regular distribution, indicating a predictable structure even in large volumes of data. This law can be applied in different contexts such as data understanding, Natural Language Processing (NLP), keyword identification in text, and SEO (*Search Engine Optimization*) strategies.

Cassettari *et al*. (2015), when analyzing the *word frequency rankings* , identified distinct patterns of use: some words are very common, such as articles and prepositions, which have a structural role in the text, while others of a technical nature, such as semantics, have lower frequencies. This not only confirmed the validity of Zipf's Law in both contexts, but also highlighted significant differences in vocabulary choice between writing and speech. Although Zipf's Law is of great value to complement a bibliometric analysis, in the search for trends based on keywords, it is necessary to consider that there are limitations in the standardization of these words, as will be described in the next section.

## CHALLENGES AND LIMITATIONS

Conducting an effective bibliometric analysis involves several challenges, from the selection of the most appropriate databases to the coherent interpretation of the results. The need for specific technical skills, the ability to integrate different methods of analysis, and the importance of interdisciplinary collaboration are crucial factors to ensure the quality and reliability of the results (Da Silva; Hayashi C.; Hayashi M., 2011).

A limitation to be considered in this methodology is the lack of homogeneity in the keywords, which can result in an incomplete or distorted analysis in the identification in techniques such as thematic trends (Geisler, 2002). This is illustrated by Cassettari *et al*. (2015, p. 161-162) in a comparative study of Zipf's Law between texts and oral discourses:

> All these words mentioned appear and/or have a direct meaning with the title of the text presented. If we consider the words "book" and "books" as one, they add up to 41 occurrences, raising them to the eighth position and, in the same case, with the words "citation" and "citations", together they would add up to 39 occurrences, moving to the ninth position (Cassetari *et al*. 2015, p. 161-162).

In the aforementioned study by Cassettari *et al*. (2015), it can be seen, therefore, that the grouping of words with the same meaning (differentiated by singular and plural) modifies the position of the terms in relation to the frequency of occurrence in a text or discourse. Thus, although the similarity between the words is evident to the researcher, it can complexify the process and lead to a fragmented analysis of the data, potentially distorting the final results of a keyword analysis. Ignoring the presence of some terms that

Revista ARACÊ
ISSN: 2358-2472

reflect the same meaning or interpreting a term with prior assumptions can result in biased analysis. One way to mitigate this is to complement the authors' keyword analysis with a keyword analysis of the titles, expanding the capacity for inferences and interpretations of the data and contributing to the understanding of emerging search trends (Chen *et al.*, 2024).

Another challenge concerns data quality. To illustrate this, the example of the simulated search in the Scopus database, mentioned in the methodology, is used. The initial result returned 5184 articles, reduced to 3331 (*dataset 1*) after applying filters and removing seven duplicates. After that, an exploratory analysis of the data was carried out, identifying flaws in the completeness of metadata, such as keywords. It was decided to intentionally exclude articles with incomplete metadata (1079), to illustrate the impact of this exclusion on the analysis, resulting in the *final dataset* (*dataset 2*), consisting of 2252 articles.

The analysis of the *datasets* revealed discrepancies in the completeness of the metadata, directly impacting the quality of the results. In *dataset 1*, the *author's keywords* field showed an absence of 21.22%, classified by *bibliometrix* as poor. This flaw compromises the accuracy of a keyword-dependent analysis. Despite this, *dataset 1* could be considered suitable exclusively for analyzing performance, since the performance-related metrics were complete. After the deletion of 1079 articles, *dataset 2* showed an improvement in metadata that was with low completeness, and they were then classified by the *software* as excellent. However, there were still flaws in the number of references cited, which could compromise an analysis of citations, in addition to the fact that a bias was introduced in the selection. These discrepancies highlight the importance of a careful choice of metrics and techniques in a bibliometric analysis. These results corroborate the criticisms identified in the literature about the limitations of bibliometric analysis, due to its dependence on accurate data, manipulation of citations, and limitations in capturing qualitative nuances of knowledge. To overcome these limitations, researchers have recommended the adoption of hybrid methods that integrate qualitative and quantitative analyses, the improvement of metadata accuracy, and the use of innovative approaches, such as webometrics and altimetry (Glänzel, 2003; Aria; Cuccurullo, 2017; Ball, 2017; Ninkov; Frank; Maggio, 2022). In addition, it is necessary to consider the possibilities of filling in this data, as will be explained below in the *proposed framework*.

Revista ARACÊ
ISSN: 2358-2472

***METHODOLOGICAL FRAMEWORK* FOR CONDUCTING A BIBLIOMETRIC ANALYSIS**

Next, with the contribution of the concepts and definitions presented, the steps used for the development of a bibliometric analysis are described. This methodological framework was built based on the literature and, mainly, on the perceptions and practical experiences of the authors. Its structure consists of 12 steps (**Chart 4**) distributed in three modules, which:

I. **Initial procedures and understanding of the data:** Includes the design of the study, definition of objectives, selection of databases, creation of search *queries* and choice of *software* and databases to be used (steps 1 to 4).
II. **Simulation:** Involves exploratory testing of *the queries* in the databases, importing and preliminary analysis of the data, creating scenarios and hypothetical questions, and identifying flaws in the metadata, ensuring that the data meets the research objectives (steps 5 to 7).
III. **Conducting bibliometric analysis and documentation:** Refers to the main stage of the study, where the sample is refined, metrics and techniques are selected based on data quality, information is visualized in the *software* interface, and analyzed and described the results (steps 8 to 12).

Revista ARACÊ
ISSN: 2358-2472

Chart 4 – *Methodological framework* with 12 steps for conducting bibliometric analysis

| Module | Step | Description and Fundamentals |
|---|---|---|
| **I Initial procedures and understanding of the data** | 1 - Definition of objectives and scope | Clearly establish the objectives of the research, elaborate the problem-questions that will guide the research and delimit the scope of the study. Based on Aria and Cuccurullo (2017), the general questions of a study can be divided into: **(i)** identification of the knowledge base, through intellectual mapping (performance metrics, frequency of co-citation); **(ii)** examination of the research frontier, to identify emerging research trends (thematic evolution, thematic map, CT); and **(iii)** production of a network structure (co-authorship analysis, co-citation). Examples of questions are: **(1)** Which authors, institutions and countries are the most productive? **(2)** What is the most searched topic? (Öztürk; Kocaman; Kanbach, 2024). |
| | 2 - Creation of the *search* query(s) | Create the *search query(s)* using consolidated descriptors or thesaurian terms, such as *MeSH, Engineering main heading*, *Engineering uncontrolled terms*, *EMTREE medical terms*, in order to minimize bias. However, the inclusion of free terms may broaden the scope of the sample, as not all published articles use consolidated descriptors or controlled terms from thesauri. This discussion should be held before the polls. You can apply Boolean operators (e.g., "*AND NOT*") to exclude irrelevant results, ensuring accuracy and specificity in data retrieval. Karger and Kureljusic (2023) developed a search *string* on cancer diagnosis by AI, consisting of two parts: one focused on technical terms, such as ("*artificial intelligence*" *AND* "*machine intelligence*"), and another related to the application domain, such as ("*cancer detect*" *AND* "*cancer diagnos**"). This approach can be adapted to the construction of *queries* focused on different perspectives, such as different thematic categories ("sustainability" *AND* "circular economy"), varied geographical contexts ("public policies" *AND* "Latin America"), interactions between concepts (ethics *AND* artificial intelligence) or even types of studies and methodologies ("bibliometric analysis" *AND* education). The collaboration of a specialist in librarianship and information science is strongly recommended (Da Silva; Hayashi C.; Hayashi M., 2011). To elaborate a search *query*, see Bramer *et al.* (2018). |
| | 3 - Choice of the analysis tool | Select a *suitable software* for data processing and bibliometric analysis. This step should be carried out before starting the searches, in order to avoid collecting data from incompatible databases or that have limitations in analysis by the *software*, such as the combination of data from different databases. Examples of tools include VOSviewer, Gephi, CiteSpace, and *bibliometrix* (Aria; Cuccurullo, 2017). The choice depends on the researcher's preference, based on his or her skills and competencies (Da Silva; Hayashi C.; Hayashi M., 2011) and from the compatibility of data types to the most suitable applications (e.g., VOSviewer for network visualization and CiteSpace for citation analysis). The *bibliometrix* (Aria; Cuccurullo, 2017) can be used to conduct a complete bibliometric analysis; however, more than one analysis tool can be used, depending on the complexity of the study, as described by Karger and Kureljusic (2023). |
| | 4 - Database selection and testing | Identify and test suitable databases for collecting articles, ensuring that sources are reliable and comprehensive. Adapt the search *query* to different databases (Bramer *et al*., 2018) and consider whether joining data from multiple databases is necessary. Regarding the joining of databases, Donthu *et al.* (2021) and Öztürk, Kocaman and Kanbach (2024) suggest working with only one base, while Echchakoui (2020) suggests (in the general context) that Scopus and WOS be used. In practice, severe limitations were identified in the joining of the Scopus, |

Revista ARACÊ
ISSN: 2358-2472

| | | |
|---|---|---|
| | | WOS, PubMed, Cochrane and Dimensions databases, using the *bibliometrix tool*. The combination of these databases resulted in problems in the completeness of metadata and the creation of random rows in the resulting dataset, making it impossible to visualize the data graphically, without proper treatment. This is because each database exports the data with a *different tag* (label or data column title), and the appropriate basis for use in *bibliometrix* is Clarivate's tags. To learn about the characteristics of different databases, see Martinez *et al.* (2023). An alternative to bring together the Scopus and WOS data is presented by Lim, Kumar and Donthu (2024). |
| **II Simulation** | 5 - Preliminary search, pre-screening and export of data | Perform searches in the chosen databases, according to the scope and previously defined filters, exporting the data from the selected database. The data files must be exported containing all possible information. In addition, it is recommended that they be generated in files with different extensions, such as CSV, BIB, RIS and XLSX, as they can be useful in eventual complementary analysis by other *software*. Donthu *et al*. (2021) suggest that at least 500 studies in the area of interest should be evaluated. However, in practice this varies, as reported in the discussion of the article. It is necessary to consider that different types of documents (articles, chapters, etc.) have different citation weights (Wallin, 2005). In this case, it is suggested: **(i)** selection of different types of documents for studies in order to understand the general dispersion of production; and **(ii)** selection of articles only for a trend study, with practical and current implications. |
| | 6 - Import, processing, exploratory analysis and validation of data | Import, process, and clean the collected data, removing duplicates and inconsistencies to ensure the integrity of the analysis. Conduct further exploratory analyses to ensure that the data are adequate. Create questions and simulate various scenarios, applying performance analysis and scientific mapping techniques, verifying the completeness of the metadata and whether the sample is sufficient to answer the survey questions. For a historical analysis, it is necessary to investigate from the first publication of the sample, without time frames. To identify recent trends, you can focus on the period of greatest increase in publications. A complete bibliometric analysis requires two approaches: **(i)** general, without time frame, and **(ii)** focused on the most current period, considering the beginning of the exponential increase in production, identified in the preliminary visualization of the data. However, sub-analyses can be performed with specific periods. Regarding the failure to complete metadata, there is the possibility of filling in the missing data manually; However, this can be a great challenge and take time, as there may be ambiguity in the data or even not be located. This is a choice of the authors, who should consider the response time necessary for the results of the study, since bibliometric analysis is a methodology with an agile profile. The automation of data processing can be done via APIs (*Application Programming Interfaces*) of the databases or custom *scripts*, and it is suggested to always consult the manual of the chosen *software* to check the possibilities. |
| | 7 - Review of the *search query(s)* (optional) | If inconsistencies or problems with data quality are found in a way that could compromise the answer to the search problem, you should adjust the search *query* and repeat the process from step 5. This feedback loop will ensure that the data meets the objectives of the survey. If problems persist, it may be necessary to revise the study objectives and restart the process from step 1. If you do not need to review the *query*, proceed to step 8. |
| **III Conducting bibliometric** | 8 - Selection of articles by double reading or double | Create a script to further refine the selection of articles based on reading the abstracts, with questions such as: **1)** Is the study design adequate to answer the proposed research problem? **2)** Have the objectives of the |

Revista
ARACÊ
ISSN: 2358-2472

| | | |
|---|---|---|
| **analysis and documentation** | blind | study been stated? **3)** Does the study adhere to the research problem [state the problem]? For greater rigor, it is suggested that the abstracts of the articles be: **(i)** selected by a specialist in the area with double reading of the abstracts; or **(ii)** selected by two independent authors (double-blind), with the support of an expert, to resolve any divergences in the selection. This procedure confers greater methodological rigor to the study, considering that, even with efforts to carefully select articles through a well-prepared *query*, related texts can be retrieved, but which do not directly adhere to the research problem. After selecting the articles, step 6 (and 7, if necessary) should be redone before proceeding to step 9. This is due to the fact that, depending on the number of articles excluded, two situations may occur: **(i)** the sample may become too small, compromising its representativeness; **(ii)** the exclusion of articles may impact the sample metrics, requiring a new analysis of the quality of the data. It is suggested that the deletion be done directly in a copy of the file (spreadsheet), to be imported later for analysis. |
| | 9 - Selection of metrics and techniques | Choose the metrics and techniques that will be applied, in an appropriate way for the bibliometric analysis, taking into account the characteristics of the available metadata, and the results identified in step 6. It will not always be necessary to conduct a complete bibliometric analysis, and it is suggested to focus on the application of techniques that will respond to research problems, according to examples and references in **tables 2** and **3**, and **table 1**. Additionally, it is recommended to consult: Byl *et al*. (2024) and Rahman *et al*. (2024). |
| | 10 - Importing data, conducting bibliometric analysis and description of results | Import the data resulting from step 8 into the chosen *software* and conduct the bibliometric analysis (Zupic; Čater, 2015; Aria; Cuccurullo, 2017; Donthu *et al*., 2021; Öztürk; Kocaman; Kanbach, 2024). After data analysis, it is recommended to use the BIBLIO guide as a basis to structure the bibliometric analysis article (Montazeri *et al.*, 2023). A sequence for conducting and describing the results, based on Donthu *et al*. (2021), involves three steps: **(i) Descriptive analysis:** presenting the general data of the sample in tables and describing them (**table 1**). **(ii) Performance analysis:** apply metrics such as TP, TC and *h-index,* and techniques such as Bradford's Law and Lotka's Law. The results can be presented in the following order: oldest article and most cited articles in the sample; most productive and most cited (influential) authors, journals, institutions and countries. **(iii) Scientific mapping:** Analysis of frequent keywords, *trend topics*, thematic mapping, and maps of international collaborations may be sufficient to understand the current state of research and emerging trends. This approach can be complemented by techniques such as bibliographic coupling, co-occurrence of keywords and factor analysis, according to the objectives. The presentation should be descriptive and results-focused, with no inferences from the authors. Aria and Cuccurullo (2017) suggest using visualizations such as networks, maps, and diagrams to facilitate interpretation and connections between data. An integrative framework (Öztürk; Kocaman; Kanbach, 2024) can organize categories of metrics, highlighting, for example, the three most productive journals, the most cited, collaboration networks, among others. This strategy guarantees essential information about scientific performance, even when the focus of the study is not performance analysis. Volpe *et al*. (2023) present the results correlating collaborations and impacts, exemplifying the flexibility in the organization of results. |
| | 11 - Analysis, interpretation of data and discussion | Discuss the study with an emphasis on bibliometric findings, drawing inferences and relationships between metrics and answering initial research questions. It is recommended that the discussion follow the order in which the data was presented. Suggested steps for the |

Revista ARACÊ
ISSN: 2358-2472

| | | |
|---|---|---|
| | | discussion of the findings, based on Montazeri *et al*. (2023) and European Radiology (2024) and their perspectives: start with a brief resumption and contextualization of the study's objective, followed by the discussion of the main metrics and bibliometric findings, and trend topics. Perform a critical analysis, confronting the results with the literature, and conclude by addressing the implications of the study (academic, social, clinical, etc.), potentialities, limitations, and future perspectives, followed by the conclusion section. It is recommended to read Öztürk, Kocaman and Kanbach (2024) to broaden the scope of the discussion. Although the focus is on quantitative analysis, a qualitative analysis of selected studies, such as the most cited or the main ones in a given *cluster*, can complement the research perspectives. |
| | 12 - Preparation of supplementary material and reproducibility manual (optional) | Bibliometric analyses are studies that often generate a large volume of data due to the complexity of the procedures and the various possibilities of analysis. The use of supplementary material should be seen as a complement to the research, preventing the main text from depending on it to make sense to the reader. The main text must be self-sufficient and clear, while supplementary material may contain additional information, such as: detailed descriptions of the search procedures, a glossary with the terms used, as well as complementary tables and figures that are not essential for understanding the study. In the supplementary material, it is also possible to detail all the methodological procedures, including the *scripts* used in the *software*, which facilitates reproducibility or adaptations by other researchers. However, this step may be waived if the *template* of the article allows a sufficiently detailed description for the reproduction of the research directly in the main text, or if the rules of the chosen journal do not support the inclusion of this type of material. To host and make available the supplementary material, it is suggested to use repositories such as Figshare (https://figshare.com), which allows the storage and sharing of data in a practical and accessible way. |

**Note:** TC = Total citations; TP = Total publications. **Source:** Prepared by the authors (2024), based on the literature review presented and practical perceptions of the application of the *framework* in other ongoing research.

## DISCUSSION

Bibliometric analysis has been consolidated as an indispensable tool to understand the quantitative panorama of scientific production in different areas of knowledge. This article presented a review of the main concepts, approaches, and challenges related to the application of this methodology. In addition, it proposed a methodological framework in 12 steps, distributed in three modules, with the objective of guiding researchers in their investigations. From this proposal, crucial elements that impact the effectiveness and quality of the analyses were identified, as well as the methodological decisions that accompany them.

### SCOPE OF LITERATURE AND DATABASES

According to Donthu *et al*. (2021), bibliometric analysis is recommended for research fields with 500 or more publications, ensuring a robust basis for metrics and mappings. For

Revista ARACÊ
ISSN: 2358-2472

the authors, in areas with few articles (50 to 300), its application is unnecessary, as significant results would hardly be obtained, and systematic or traditional reviews would be more appropriate. However, in the literature, studies that covered variable sample sizes were identified, such as 6450 studies (Karger; Kureljusic, 2023), 31169 (Ruiz-Fresneda; Morales-Álvarez, 2024) and 164 articles (Ínan, 2023).

Two points deserve to be highlighted. The first is that, for emerging and recent themes, the sample of studies may not be broad enough, but the bibliometric study can contribute to identifying the preliminary bases of the literature. Periodic repetition of the analysis (for example, every one or two years) can provide an up-to-date follow-up of research, keeping the science updated on a given topic. The second point refers to the fact that, although a sample of thousands of documents is used, there may be a lack of homogeneity in the weight of citations (Wallin, 2005), especially if it covers different types of documents (such as articles and book chapters). In addition, potentially distorted results may arise if articles are not properly screened for adherence to the theme and the analysis is conducted on a sample generated only based on automatic filters. To mitigate this problem, the proposed methodological framework suggests the inclusion of a step (8) for manual screening, with double checking of the text, or, alternatively, by two independent reviewers, with the help of an expert, ensuring that the analysis is conducted on a consistent sample, removing articles that are impertinent for the study.

Regarding the selection of databases, there is an excessive dependence on the tool to be used. The merger of data sets can come up against practical aspects, due to the divergence in their structure, according to each database. In the literature, studies based mainly on Scopus and the Web of Science (WOS) have been identified, although other databases such as Medline, Philosophers Index and Lilacs have also been used (Pizzani *et al*., 2010). Donthu *et al*. (2021) and Öztürk, Kocaman and Kanbach (2024) suggest the use of only one database. On the other hand, Echchakoui (2020) points out that the merger of Scopus and WOS can bring more consistency to a research. In the present practical tests, the combination of the Scopus, WOS, Dimensions, PubMed and Cochrane databases, supported by *bibliometrix*, resulted in problems in the datasets, such as the insertion of blank lines, which made it impossible to import the file into the graphical interface by the *biblioshiny()* function, without them being deleted directly in the spreadsheet. It also resulted in metrics that, although complete initially, became incomplete, which possibly occurred due to the different nomenclatures of the *tags* in each database.

Revista
ARACÊ
ISSN: 2358-2472

Lim, Kumar, and Donthu (2024) suggest an alternative to combine the Scopus and WOS bases. Currently, *bibliometrix* has a function to merge datasets in the graphical interface. It is always recommended to confirm the real need to join the databases, as most journals indexed in WOS are also indexed in Scopus. In addition, the need to pay attention to the final quality of the metadata is reinforced.

In a 24-step guide to conducting a systematic review, Muka *et al*. (2019, p. 52) state that "a bibliographic search must include at least four online databases: Embase, MEDLINE, Web of Science, and Google Scholar". Although the interest of the systematic review is to assess the quality of the evidence, this raises a question about the scope and representativeness of the sample. Therefore, it is reflected that working exclusively with databases such as WOS and Scopus, although popular, may limit the representativeness of the sample, since potential studies for the analysis may be contained in other databases.

## ANALYSIS STRUCTURE AND METRICS

Bibliometric analyses are traditionally structured on three main fronts: descriptive, performance, and scientific mapping (Donthu *et al.* 2021). This methodological balance allows not only quantitative interpretation, but also indirect qualitative contextualization of the identified interrelationships (Geisler, 2000; Glänzel, 2003). Despite its quantitative nature, integration with a brief qualitative synthesis or analysis—such as the most cited or most relevant studies within a given *cluster*—can enrich the interpretation of results. This approach expands the relevance of scholarly contributions, providing a more comprehensive view of trends, gaps, and potential practical applications of bibliometric findings. This integration is reaffirmed by the study by De Medeiros Filho and Russo (2018), on brands as an indicator in the context of companies, which mixed a systematic review with bibliometric analysis, demonstrating how methodologies can complement each other to offer broader *insights*.

## IMPACT OF METADATA QUALITY

Wallin (2005) points out that the absence of metadata, such as authors' full addresses in 17% of the SCIE (*Science Citation Index Expanded*) records, compromises bibliometric analyses, especially at intermediate levels, making it difficult to identify institutional affiliations and assess scientific impact. This tends to be aggravated by the lack of standardization of the names of the institutions. On the other hand, when trying to fill in manually, the researcher may be faced with the multiple and simultaneous affiliation of an author, which can compromise the accuracy of the bibliometric analysis by distorting metrics such as productivity, collaboration, and academic impact.

The example of the exclusion of articles with keyword failures in the *analyzed datasets* (**Table 1**) illustrates how the absence of data can compromise analyses dependent on the co-occurrence of terms, reducing the amount of information available. However, it should be noted that additional keywords cannot be corrected manually, as they are automatically generated by an algorithm from the titles of the references cited by an article, which limits the researcher's intervention and may introduce biases in the results of the analysis.

It is worth noting that the exclusion of articles with metadata flaws should be conducted with the utmost caution, as it can significantly impact and alter the results of the metrics, in addition to introducing bias, weakening the methodological bases of the research. Thus, any deliberate decision to work with limitations in the completeness of metadata or to exclude studies should be discussed with an expert in the field and duly reported in the methodology or limitations section of the study, alerting readers to the potential risks in interpreting the results.

It is expected that a bibliometric analysis will produce results in an agile way, considering the frequent insertion of new publications in the literature, which can impact the results on emerging topics. This suggests that, in innovation topics, there may be a rapid obsolescence of bibliometric analysis, which makes it necessary to carry out new research periodically on a topic, so that decisions are based on updated and relevant data.

## POTENTIALITIES, LIMITATIONS AND FUTURE PROSPECTS

The methodological framework presented demonstrates that flexibility in the choice of metrics and techniques is essential to adapt the analysis to the specific objectives of the research. In addition, it can be adjusted to conduct research in any area of knowledge. Simulated tests and *feedback* loops, intentionally inserted, contribute to adjusting the focus of the investigation, ensuring that the sample is representative and adequate to the problem questions, in addition to providing more transparent and reliable results.

The main limitation is that only *bibliometrix* was used to conduct the analyses, and it is suggested that other tools be added to broaden the perspective of the *framework* and explore different applications of techniques to the data. As a continuity, it is suggested the

Revista ARACÊ
ISSN: 2358-2472

adoption of multimodal approaches, combining quantitative and qualitative elements to achieve a more comprehensive and detailed understanding of scientific production. It is also proposed to apply the *framework* in different databases and thematic contexts, in order to refine and expand the model presented. Finally, the need to develop strategies to maintain rigor in the completeness of metadata and in the integration between different databases is emphasized, enabling a truly broad and representative analysis of the literature.

## CONCLUSION

The present study presented a narrative review on the methodology of bibliometric analysis, highlighting its importance as a tool for quantitative evaluation and the connections that are formed in scientific production, and proposed a methodological framework in 12 steps, to be used by researchers and students interested in the subject. Bibliometric analysis, despite its limitations, remains an important methodology to map the productivity, influence, and communities that form around academic-scientific production, as well as to identify emerging research themes. The application of the proposed *framework* can contribute to more rigorous and reliable methodological practices. As a continuity, it is suggested to explore the integration of new metrics, hybrid approaches (quantitative and qualitative) and advanced analysis techniques, expanding the applicability and depth of bibliometric analyses.

## ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Revista ARACÊ
ISSN: 2358-2472

## REFERENCES

1. ABBAS, A. F. et al. (2022). Bibliometrix analysis of information sharing in social media. Cogent Business & Management, 9(1), 2016556.

2. ALABI, G. (1979). Bradford's law and its application. International Library Review, 11(1), 151-158.

3 ALVARENGA, L. (1998). Bibliometria e arqueologia do saber de Michel Foucault: traços de identidade teórico-metodológica. Ciência da Informação, 27(3), set.

4. ARIA, M. & CUCCURULLO, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959-975.

5. BALL, R. (2017). An Introduction to Bibliometrics. Amsterdam: Chandos Publishing.

6. BRAMER, W. M. et al. (2018). A systematic approach to searching: an efficient and complete method to develop literature searches. Journal of the Medical Library Association: JMLA, 106(4), 531.

7. BYL, L. et al. (2016). Measuring Research Output through Bibliometrics. Disponível em: https://bit.ly/3YbJujs. Acesso em: 16 out. 2024.

8. CASSETTARI, R. R. B. et al. (2015). Comparação da Lei de Zipf em conteúdos textuais e discursos orais. El profesional de la información, 24(2), 157-167.

9. CHEN, G. et al. (2024). Comparing semantic representation methods for keyword analysis in bibliometric research. Journal of Informetrics, 18(3), 101529.

10. DA SILVA, M. R.; HAYASHI, C. R.; HAYASHI, M. C. P. I. (2011). Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. InCID: revista de ciência da informação e documentação, 2(1), 110-129, 9 jun.

11. DONTHU, N. et al. (2021). How to conduct a bibliometric analysis: an overview and guidelines. Journal of Business Research, 133, 285-296.

12. ECHCHAKOUI, S. (2020). Why and how to merge Scopus and Web of Science during bibliometric analysis: the case of sales force literature from 1912 to 2019. Journal of Marketing Analytics, 8, 165-184.

13. EGGHE, L. (2006). Theory and practice of the g-index. Scientometrics, 69, 131-152, 20 jun.

14. ELSEVIER. (2009). Research Metrics Guidebook. Disponível em: https://bit.ly/4b9PfnH. Acesso em: 07 ago. 2024.

15. EUROPEAN RADIOLOGY. (2024). European Radiology Manuscript requirements for submissions (Ed. 2). Disponível em: https://bit.ly/3CNn54V. Acesso em: 24 nov. 2024.

16. FERRARI, R. (2015). Writing narrative style literature reviews. Medical Writing, 24(4), 230-235.

17. GEISLER, E. (2000). The metrics of science and technology: evaluation and measurement of research, development, and innovation. Westport, CT: Quorum Books.

18. GLÄNZEL, W. (2003). Bibliometrics as a research field: a course on theory and application of bibliometric indicators. Disponível em: https://bit.ly/3yr5OwZ. Acesso em: 14 ago. 2024.

19. GREEN, B. N.; JOHNSON, C. D.; ADAMS, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. Journal of Chiropractic Medicine, 5(3), 101-117.

20. GUIMARÃES, A. J. R.; MOREIRA, P. S. C.; BEZERRA, C. A. (2021). Modelos de inovação: análise bibliométrica da produção científica. Brazilian Journal of Information Science, 15, 6, 4 maio.

21. HIRSCH, J. E. (2005). An index to quantify an individual's scientific research output. arXiv:physics/0508025 [physics.soc-ph], 102(46), 16569-16572, 29 set.

22. HJØRLAND, B. (2004). Domain analysis: A socio-cognitive orientation for information science research. Bulletin of the American Society for Information Science and Technology, 30(3), 17-21, 31 fev.-mar.

23. İNAN, Ü. S. E. (2023). Evaluation of digital marketing from a bibliometric analysis perspective. Socialis Series in Social Science, 4, 45-58.

24. KARGER, E.; KURELJUSIC, M. (2023). Artificial intelligence for cancer detection—a bibliometric analysis and avenues for future research. Current Oncology, 30(2), 1626-1647.

25. LIM, W. M.; KUMAR, S.; DONTHU, N. (2024). How to combine and clean bibliometric data and use bibliometric tools synergistically: guidelines using metaverse research. Journal of Business Research, 182, 114760.

26. LOTKA, A. J. (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16(12), 317-323, 19 jun.

27. MARTINEZ, E. C. et al. (2023). Ten steps to conduct a systematic review. Cureus, 15(12).

28. MEDEIROS FILHO, A. R.; RUSSO, S. L. (2018). Marcas como um indicador: revisão sistemática e análise bibliométrica da literatura. Biblios, 71, 50-67.

29. MILES, M. B.; HUBERMAN, A. M. (1994). Qualitative data analysis: an expanded sourcebook (2nd ed.). Thousand Oaks: SAGE Publications.

30. MOMESSO, A. C.; NORONHA, D. P. (2017). Bibliométrie ou Bibliometrics: o que há por trás de um termo? Perspectivas em Ciência da Informação, 22, 118-124.

Revista ARACÊ
ISSN: 2358-2472

31. MONTAZERI, et al. (2023). Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. Systematic Reviews, 12(1), 15 dez.

32. MUKA, T. et al. (2020). A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. European Journal of Epidemiology, 35, 49-60.

33. MUKERJEE, S. (2021). A systematic comparison of community detection algorithms for measuring selective exposure in co-exposure networks. Scientific Reports, 11(1), 15218.

34. NAVARRO-BALLESTER, A. et al. (2023). Publications on COVID-19 in radiology journals in 2020 and 2021: bibliometric citation and co-citation network analysis. European Radiology, 33(5), 3103-3114, maio.

35. NEWMAN, Mark E. J. (2010). Networks: an introduction. Oxford; New York: Oxford University Press.

36. NINKOV, A.; FRANK, J. R.; MAGGIO, L. A. (2022). Bibliometrics: Methods for studying academic publishing. Perspectives on Medical Education, 11(3), 173-176.

37. N'ZE, A. A. P.; TENKOUL, A. (2024). Global Research on Good Governance and Sustainability: a bibliometric analysis. International Journal of Sustainability in Economic, Social, and Cultural Context, 20(2), 47-76, 20 set.

38. ÖZTÜRK, O.; KOCAMAN, R.; KANBACH, D. K. (2024). How to design bibliometric research: an overview and a framework proposal. Review of Managerial Science, 1-29, 6 mar.

39. PASSAS, I. (2024). Bibliometric Analysis: The Main Steps. Encyclopedia, 4(2), 1014–1025.

40. PIZZANI, L.; DA SILVA, R. C.; HOSSNE, W. S. (2010). Análise bibliométrica dos 40 anos da produção científica no Brasil e no mundo. Revista Bioethikos, 4(4), 453-460.

41. POSIT. (2024). RSTUDIO IDE The most trusted IDE for open source data science. Disponível em: https://posit.co/products/open-source/rstudio/. Acesso em: 06 out. 2024.

42. RAHMAN, A. et al. (2024). A comprehensive bibliometric and content analysis of artificial intelligence in language learning: tracing between the years 2017 and 2023. Artificial Intelligence Review, 57(4), 107.

43. RUIZ-FRESNEDA, M. A.; GIJÓN, A.; MORALES-ÁLVAREZ, P. (2023). Bibliometric analysis of the global scientific production on machine learning applied to different cancer types. Environmental Science and Pollution Research, 30(42), 96125-96137, set.

Revista ARACÊ
ISSN: 2358-2472

44. SMITH, N. R. et al. (2020). A guide for choosing community detection algorithms in social network studies: the question alignment approach. American Journal of Preventive Medicine, 59(4), 597-605.

45. TAGUE-SUTCLIFFE, J. (1992). An introduction to informetrics. Information Processing & Management, 28(1), 1-3.

46. TERRA, et al. (2022). Análise bibliométrica com o software bibliometrix. In: XLII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2022, Foz do Iguaçu, PR. Anais [online]. Foz do Iguaçu: Enegp. Disponível em: https://bit.ly/4cyl1Ko. Acesso em: 20 ago. 2024.

47. URIBE, F. F. M.; GUERRERO, J. C. O.; CONTRERAS, F. C. (2023). Introducción a la bibliometría práctica. Asociación de Bibliotecólogos del Perú. Disponível em: https://bit.ly/3WWNncN. Acesso em: 10 ago. 2024.

48. VAN NOORDEN, R. (2010). A profusion of measures. Nature, 465(7300), 864-867, jun.

49. VASCONCELOS, Y. L. (2014). Estudos bibliométricos: procedimentos metodológicos e contribuições. Revista de Ciências Jurídicas e Empresariais, 15(2), 211-220, set.

50. VOLPE, S. et al. (2023). Quo vadis Radiomics? Bibliometric analysis of 10-year Radiomics journey. European Radiology, 33(10), 6736-6745.

51. WALLIN, J. A. (2005). Bibliometric methods: pitfalls and possibilities. Basic & Clinical Pharmacology & Toxicology, 97(5), 261-275.

52. ZHANG, Y. et al. (2024). Artificial intelligence in sign language recognition: a comprehensive bibliometric and visual analysis. Computers and Electrical Engineering, 120, 109854.

53. ZIPF, G. K. (1949). On the Economy of Words. In ZIPF, G. K. Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology. Massachusetts: Addison-Wesley Press, INC, 19-55. Disponível em: https://bit.ly/3OsYbKO. Acesso em: 24 nov. 2024.

54. ZUPIC, I.; ČATER, T. (2015). Bibliometric methods in management and organization. Organizational Research Methods, 18(3), 429-472.