Revista PANAMERICANA de COMUNICACIÓN

Vol. 7, n. 1, enero - junio 2025

A critique of 'quick and dirty' bibliometrics: issues for assessing complexity in bibliometric analysis

Contra la bibliometría "rápida y sucia": aspectos para valorar la complejidad en los análisis bibliométricos

Rafael Repiso; Álvaro Cabezas-Clavijo

Rafael Repiso

Universidad de Granada, España
https://orcid.org/0000-0002-2803-7505
repiso@ugr.es

Álvaro Cabezas-Clavijo

Universidad Internacional de La Rioja, España
https://orcid.org/0000-0001-9641-8855
alvaro.cabezas@unir.net

How to cite this article

Repiso, R., & Cabezas-Clavijo, Á. (2025). Contra la bibliometría "rápida y sucia": aspectos para valorar la complejidad en los análisis bibliométricos. *Revista Panamericana de Comunicación*, *7*(1), 3419. https://doi.org/10.21555/rpc.v7i1.3419

Received: 22 - 04 - 2025
Acepted: 05 - 05 - 2025
Published online: 14-06-2025

Abstract

This paper denounces the indiscriminate rise of low-quality bibliometric studies. These studies have proliferated across scientific journals in a wide range of disciplines, taking advantage of the accessibility of databases such as Web of Science and Scopus and the use of automated tools. This trend, driven by a lack of specialised knowledge among journal reviewers and a desire to rapidly increase academic output, has led to the publication of studies with methodological shortcomings, characterised by limited conceptual development and low analytical value. In response to this situation, the authors advocate for bibliometrics as a highly specialised field that requires theoretical understanding, methodological expertise, and critical interpretation of data. Thus, the paper proposes a methodological framework for assessing the complexity of bibliometric studies. This framework is based on six dimensions: size of the population, origin and source of data, data collection mode, degree of data normalisation, types of analysis employed, and analytical and visualisation tools used. This methodological perspective aims to provide editors and researchers with a framework for identifying substantive research and distinguishing it from studies produced with least effort, limited judgment, and weak theoretical grounding.

Keywords: Bibliometrics; Methodology; Quality; Scientific databases: Data analysis; Data visualisation.

Resumen

El presente trabajo denuncia el auge indiscriminado de estudios bibliométricos de baja calidad que, aprovechando la accesibilidad de bases de datos como Web of Science o Scopus y el uso de herramientas automatizadas, proliferan en

revistas científicas de diversas disciplinas. Esta tendencia, facilitada por la ausencia de conocimientos especializados de los revisores de revista y por el interés en incrementar rápidamente la productividad académica, ha derivado en la publicación de investigaciones metodológicamente pobres, con escasa elaboración conceptual y valor analítico limitado. Frente a este panorama, los autores reivindican la bibliometría como un campo de alta especialización que exige comprensión teórica, competencia metodológica e interpretación crítica de los datos. Con base en estos principios, se propone una aproximación metodológica para valorar la complejidad de los estudios bibliométricos a partir de seis dimensiones: tamaño de la población, origen y fuente de los datos, forma de recolección, grado de normalización, tipos de análisis empleados y herramientas utilizadas. Esta perspectiva metodológica pretende ofrecer a editores e investigadores un marco para identificar investigaciones sustantivas y distinguirlas de aquellas realizadas con poco esfuerzo, criterio y contexto teórico.

Keywords: Bibliometría; Metodología; Calidad; Bases de datos científicas: Análisis de datos; Visualización de datos.

The plague of bibliometric analyses

A plague of bibliometric studies is sweeping across scientific journals of all disciplines and calibres. While the number of experts in the field has not grown significantly, the number of works produced by non-experts, sometimes even commissioned, has increased exponentially, making bibliometrics fashionable in recent years. This surge in bibliometric output can largely be attributed to two key factors. First, the ease of data collection. Major scientific databases, primarily Web of Science and Scopus, now allow users to retrieve and organise 'clean' data in minutes, unlike other research techniques which can take days, weeks or even years. Secondly, basic descriptive analysis (such as author, journal or institutional counts) is now automatically provided by the databases themselves, while software such as VOSviewer or Bibliometrix enables virtually anyone —with no specific training in bibliometrics — to generate indicators and visually appealing outputs for a general audience.

We are witnessing a surge in the publication of 'easy' bibliometric works: pieces of writing with minimal academic merit lacking analytical depth and contextual or theoretical foundation. There is little intellectual merit in appropriating data processed with little to no human intervention by software. In other words, a research paper whose data retrieval and analysis can be replicated in under a day—often in less than an hour—should not be considered worthy of publication in a serious academic journal. Such studies merely contribute to the proliferation of low-quality, low-impact science and increase the amount of noise in the scholarly ecosystem. Who would turn down an easy and quick boost to their publication record? And who would say yes? Often, these are researchers with little or no background in scientometrics who see bibliometrics as a convenient way to increase their academic output.

This group of academic migrants is characterised, in particular, by publishing their bibliometric studies in journals unrelated to Library and Information Science, since specialised journals often reject their work outright. Consequently, they avoid submitting to journals with expert peer reviewers. Instead, they have found a profitable niche in megajournals and white-label open-access outlets of major academic publishers, which allow for fast and effortless publication. Their work usually involves repeating the same methodological process across different datasets or disciplines.

However, the field has also historically attracted a second group of researchers that Emilio Delgado termed 'bibliometric parachutists': those from quantitatively oriented disciplines, mostly in the natural sciences, who apply bibliometric techniques within their own fields. These

scholars often have a much better grasp of quantitative methods than the average information scientist. They use scientometrics as a method rather than the subject of their research, and they typically publish in top-tier journals within their primary disciplines, making substantial contributions. It is important to recall that many of the founding figures in the quantitative study of scientific communication came from physics or chemistry. The critical difference with today's opportunistic wave is that, for these pioneers, bibliometrics were a secondary area of research — a tool to support evaluative or descriptive studies of large scientific domains or even a hobby. (How else might we characterise Hirsch's proposal of the h-index?) Above all, they add meaningful contributions to the development of bibliometrics.

Far from being a mere mechanical technique that can be boiled down to automated charts, bibliometrics is a rigorous and highly specialised research domain that requires theoretical understanding, methodological expertise and the ability to critically interpret data. It has long been associated with the evaluation of scientific knowledge and the development of science, technology and innovation policies (Moed, 2005; Leydesdorff, 2001). High-quality bibliometric studies address complex issues such as the dynamics of scientific production, the structure of collaboration networks and the thematic evolution of disciplines, rather than simply counting publications (Waltman & Van Eck, 2015). In this sense, bibliometrics is a social science discipline in its own right, shedding light on behaviour and trends within the scientific community. When applied rigorously, bibliometrics can reveal hidden patterns in the scientific system and help to test hypotheses of considerable strategic value. Reducing it to a superficial descriptive exercise not only impoverishes its potential, but also distorts its essential role in fostering critical understanding of science.

The scholarly community itself is beginning to highlight the weaknesses of many bibliometric studies (Cabezas-Clavijo et al., 2023) and propose protocols and frameworks to ensure more robust, methodologically sound analyses. Examples include the Guidance List for Reporting Bibliometric Analyses (GLOBAL) (Ng et al., 2023); the BIBLIO guidelines for bibliometric reporting in biomedical research (Montazeri et al., 2023); and the VALOR framework for evaluating multi-source bibliometric studies (Hoang, 2025)..

How to assess the quality of a bibliometric article

As previously stated, the quality of a bibliometric study is defined by three fundamental dimensions: theoretical understanding, methodological proficiency and critical interpretation of data. Among these, the methodological component is arguably the most accessible and objective to assess. One could argue that theoretical understanding requires scientific knowledge on the application and conceptual basis of bibliometric techniques, as well as on the subject of the study — something that few experts possess. Interpreting results is a derivative of this theoretical comprehension and primarily materialises in the discussion and conclusion sections of the paper.

For this reason, the work presented here seeks to determine the value of a bibliometric article through a methodological lens, as this allows for a more objective evaluation when the constituent stages are identified and their execution is valued in terms of the time, knowledge and complexity involved.

It is important to acknowledge that the complexity of a study does not necessarily correlate with the significance of its findings. Likewise, a sound methodology does not necessarily guarantee impactful results. Some of the most significant contributions to the field are theoretical in nature. However, if the objective is to evaluate methodological robustness, assessing the complexity of the process becomes crucial, at least to estimate the effort involved.

So, how does one determine the value of an academic article? This is a difficult question, akin to the central question in economics: what is the value of things? This issue has been debated since Aristotle distinguished between use value and exchange value, and continues to be debated by modern theorists who argue that value is subjective and may derive from a multitude of sources. This paper takes a simplified approach based on the ideas of thinkers such as David Ricardo, Adam Smith and Karl Marx, who believed that the value of an object came from the labour required to produce it (time, materials, tools, effort, etc.).

Accordingly, this study proposes a contribution-based approach, since it is often impossible to assess other forms of value, such as scientific, economic or theoretical value, at the time of article evaluation.

In our view, the complexity of a bibliometric study is defined by six key elements:

- Corpus size: The number of records or items analysed in the study.
- Data sources and origins: Whether the data come from one or multiple databases, and whether these are open or proprietary.
- Method of data retrieval: The procedures used to extract the data and whether they are automated or manual.
- Degree of data cleaning and normalisation: The extent to which the data requires disambiguation, deduplication, name standardisation and error correction.
- Types of analyses performed: These range from basic descriptive statistics to advanced network or temporal analyses.
- Software used for data analysis and visualisation: The complexity and capabilities of the tools employed, including whether custom programming or advanced statistical methods are involved.

Corpus size

The length of the analysed document corpus directly affects the methodological design, depth of analysis and tools employed. Studies based on small samples (e.g., fewer than 100 documents or 200 in the case of ScienceDirect) typically adopt a qualitative-quantitative approach, combining bibliometric analysis with a critical content review. One type of study that has gained popularity in recent years is bibliometric analysis combined with a systematic review. However, these analyses should be approached with caution, as neither technique is often applied with the required rigour. Furthermore, the necessity of conducting a bibliometric study on such a small dataset may be questioned. When data are scarce, manual or semi-automated methods are usually employed, such as co-authorship analyses, cited references and reviews of the argumentative structure of texts.

Qualitative approaches become less viable in medium-sized corpora (e.g., between 100 and 5,000 records), but the range of bibliometric techniques expands. Co-authorship network analysis, keyword co-occurrence analysis, co-citation analysis and thematic clustering are all feasible here. In large-scale studies involving over 5,000 documents, complexity increases exponentially. Advanced data mining techniques, machine learning, multivariate analysis and natural language processing (NLP) are essential for classifying data, reducing dimensionality and detecting meaningful patterns. The volume of data also determines the necessary storage and processing infrastructure, ranging from simple spreadsheets to interconnected database systems.

Data sources and origins

Data sources are not neutral repositories of information. Each one represents different types of content with distinct inclusion criteria, selection processes, normalisation standards and authority control mechanisms. They also differ in terms of accessibility, ranging from platforms that allow the bulk downloading of fully standardised datasets, to those that permit only limited viewing or partial exports. Downloading 50,000 records can take as little as five minutes with some products, or several days with others.

Using traditional databases such as Web of Science or Scopus involves minimal complexity and remains the predominant approach in bibliometric studies. However, incorporating alternative sources—global (e.g., Dimensions, Crossref, OpenAlex) or regional (e.g., Redalyc, Dialnet, Scielo —allows for new perspectives on scientific production and more accurately reflects the bibliodiversity of the academic ecosystem.

Method of data retrieval

As mentioned, some databases offer built-in data export services, which are sometimes subject to quantitative limits. Others do not allow downloads and can make access deliberately difficult (e.g., retrieving records from a university library catalogue). Data extraction from sources such as MEDLINE or Dimensions may require familiarity with APIs or programming skills in R or Python, particularly when handling large datasets. Meanwhile, open-access platforms such as OpenAlex have democratised access to scholarly data. However, making full use of these resources also demands proficiency in complex data structures (e.g., JSON) and advanced computational processing. The variety of data retrieval methods means that bibliometric researchers must craft robust search strategies—a foundational aspect that, if flawed, can invalidate an entire study—and possess or coordinate the technical skills necessary for efficient, reliable data collection. Therefore, the chosen source and the method of retrieval are key indicators of a study's complexity and quality.

Degree of data cleaning and normalisation

The degree to which data is normalised and corrected is a critical aspect of bibliometric studies, as their validity depends heavily on the accuracy and standardisation of the data. Although products such as Web of Science and Scopus offer relatively well-structured datasets, particularly with regard to their Anglo-American coverage, inconsistencies still occur in author names, institutional affiliations, keywords and cited references. Such issues are more prevalent in databases with limited resources or in large-scale, interdisciplinary studies. Such inconsistencies can result in duplication, misattribution of documents or distortion of actual collaboration and impact patterns. Normalisation involves complex procedures such as reconciling name variants (e.g., 'Univ Autonoma Madrid' vs. 'Universidad Autónoma de Madrid'), author disambiguation and harmonisation of keywords. It also often involves manual review of fields that automated algorithms cannot accurately resolve. Although often invisible to the final reader, this step is one of the most technically and epistemologically demanding phases of bibliometric analysis. A poorly cleaned dataset can lead to erroneous or irrelevant conclusions. When using databases rarely employed in bibliometric research, it is essential to verify data consistency and allocate time for authority control and normalisation — tasks that can be automated to some extent, provided the researcher has the necessary expertise. The process becomes even more complex when combining multiple sources, as each producer structures and processes content differently.

Types of analyses performed

The type of analysis employed in bibliometric studies influences the depth of the work, its explanatory power, and consequently, its scientific value. Many studies rely solely on descriptive statistics, such as simple counts of publications, citations, authors or keywords. While these are useful for providing a preliminary overview (and are often provided directly by the database itself), they are insufficient for interpreting complex dynamics. A more refined level of analysis incorporates bivariate statistics, such as correlations between variables or comparisons across fields, countries or time periods. These allow basic yet meaningful relationships to be identified. The most sophisticated studies employ multivariate techniques, such as factor analysis, clustering or principal component analysis. These techniques open the door for the identification of latent patterns, the thematic grouping of scientific production and the exploration of complex knowledge structures. Additionally, sociometric tools derived from social network analysis are frequently employed in co-authorship, co-citation, co-word, or institutional linkage studies. These graph-based methods and structural metrics provide a relational mapping of the scientific system, moving far beyond mere counting. As a rigorous and multifaceted discipline, bibliometrics must transcend superficial metrics and uncover deeper relationships among actors, discourses and the dynamics of scientific production. This requires technical and methodological mastery to distinguish substantive research from merely illustrative reporting.

Software used for data analysis and visualisation

The choice of analytical software is not a trivial matter, but rather a methodological decision influenced by the researcher's expertise and the resources available. It also determines the types of analysis that can be performed and the level of complexity that can be achieved. For basic tasks, general-purpose tools such as spreadsheets may suffice. More specialised software includes BibExcel, Bibliometrix, Loet Leydesdorff's tools and network-oriented platforms such as the Sci2 tool, CiteSpace and VOSviewer. These tools vary widely in complexity and ease of use. Some are so user-friendly that they produce visually appealing results without requiring any specialist knowledge (e.g. VOSviewer), which can lead to analyses being performed without the necessary decision-making or methodological rigour. These are often considered 'plug-and-play' tools. The highest level of expertise is required for custom programming (e.g., in R, Python, or MATLAB) to download, analyse, and visualise data. The selection and use of tools clearly indicates the capabilities of the researcher and the methodological complexity of the study.

Conclusions

Several scientific journals have already communicated their concern about the growing number of trivial bibliometric studies and the uncritical application of bibliometric tools, which is increasingly producing results of little or no academic value (Hulland, 2024). A competent bibliometric researcher possesses a diverse "toolbox" that provides access to a wide range of documentary sources beyond the frequently used Web of Science and Scopus journal databases, as well as knowledge of statistical software and data visualisation tools. A diverse methodological repertoire invariably expands research possibilities. When practised rigorously, bibliometrics is far from being a 'quick and dirty' technique that can easily be replicated. Its truee value lies in the methodological complexity involved, from corpus selection to critical interpretation of findings. Its value as a research approach resides in uncovering data that is not readily available through a simple query and in shedding light where no tool has yet illuminated; it is not merely about emulating what is already visible.

This methodological study presents six key factors for determining the complexity of a bibliometric analysis: (1) Corpus size (2) Data sources and origins (3) Method of data retrieval (4) Degree of data cleaning and normalization (5) Types of analysis performed (6) Software for data analysis and visualisation. This framework is not intended to cover all aspects of scientific value in research, but rather to provide a conceptual foundation for distinguishing between substantive contributions and those that, despite their visual appeal and ease of execution, merely add to informational noise rather than contributing to genuine knowledge. Our aim is to enable journal editors and researchers who are unfamiliar with these techniques to assess the methodological complexity and depth of a paper, if not its definitive value. Reclaiming bibliometrics as a sp ecialised field also involves protecting the ethics of effort in an academic environment that often prioritises quantity over quality.

Authors' Contribution

Rafael Repiso: Formal analysis; Conceptualisation; Writing – original draft; Writing – review and editing.

Álvaro Cabezas-Clavijo: Análisis formal; Conceptualisation; Writing – original draft; Writing – review and editing.

References

Cabezas-Clavijo, A., Milanés-Guisado, Y., Alba-Ruiz, R., & Delgado-Vázquez, Á. M. (2023). The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics. *Journal of Data and Information Science*, *8*(4), 10-35. https://doi.org/10.2478/jdis-2023-0021

Hoang, A. D. (2025). Evaluating bibliometrics reviews: A practical guide for peer review and critical reading. *Evaluation Review*, OnlineFirst. https://doi.org/10.1177/0193841X251336839

Hulland, J. (2024). Bibliometric reviews-some guidelines. *Journal of the Academy of Marketing Science*, *52*(4), 935-938. https://doi.org/10.1007/s11747-024-01016-x

Leydesdorff, L. (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Universal Publishers.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.

Montazeri, A., Mohammadi, S., Hesari, P. M., Ghaemi, M., Riazi, H., & Sheikhi-Mobarakeh, Z. (2023). Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. *Systematic Reviews*, *12*(1), 239. https://doi.org/10.1186/s13643-023-02410-2

Ng, J. Y., Haustein, S., Ebrahimzadeh, S., Chen, C., Sabe, M., Solmi, M., & Moher, D. (2023). *Guidance List for repOrting Bibliometric AnaLyses (GLOBAL): A research protocol*. https://osf.io/nvu6w

Waltman, L., & Van-Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872–894. https://doi.org/10.1016/j.joi.2015.08.001