Chapter 9 / Capítulo 9

Applied bibliometrics. From data to publication (English Edition)
ISBN: 978-9915-9680-7-0
DOI: 10.62486/978-9915-9680-7-0.ch09

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

Other Results / Otros Resultados

Analysis of complementary dimensions in bibliometrics

Beyond conventional indicators of production and impact, bibliometric analysis encompasses essential contextual dimensions that reveal structural patterns in the scientific ecosystem. This chapter examines three dimensions that are often underestimated but critical to a comprehensive understanding of research dynamics: the linguistic distribution of scientific production, the role of journals as channels of specialized communication, and the geopolitical positioning of knowledge. Each of these dimensions provides unique perspectives on how scientific knowledge is organized, communicated, and distributed across different cultural, institutional, and geographical barriers.

9.1. Language interpretation in scientific production

Analysis of language distribution in scientific literature reveals profound asymmetries in the geopolitics of knowledge.

The absolute dominance of English as *the scientific lingua franca*, which typically accounts for between 80 % and 95 % of publications indexed in international databases, reflects not only communicative standardization practices but also consolidated academic power structures. This linguistic hegemony has significant implications for the visibility, accessibility, and impact of research produced in other languages, creating a systematic bias that marginalizes valuable scientific contributions developed in peripheral linguistic contexts. The interpretation of these patterns requires historical contextualization that considers how the expansion of English as a scientific language has been linked to processes of academic globalization and the concentration of publishing resources in English-speaking countries.[(1)]

The multilingual distribution of scientific output varies substantially across disciplines, reflecting different intellectual traditions and target audiences. In fields such as clinical medicine or environmental sciences, where research has immediate local application, it is more common to find publications in languages other than English that communicate findings relevant to specific contexts. In contrast, in disciplines such as physics or mathematics, standardization in English is almost absolute.

These disciplinary differences have important implications for scientific evaluation policies, which must recognize the legitimacy of publications in multiple languages when they respond to the specific communication needs of each field of knowledge.

The evolution of linguistic patterns over time shows contradictory trends: on the one hand, a progressive consolidation of English as the universal scientific language; on the other, periodic resurgences of publications in local languages driven by open science policies and recognition of epistemological diversity. A diachronic analysis of these trends allows us to identify moments of change in scientific communication practices, usually linked to transformations in editorial policies, the emergence of multilingual repositories, or changes in academic evaluation systems that differentially value publications in different languages.

9.2. Interpretation of the role of scientific journals

Academic journals are fundamental institutions that structure scientific fields through their functions of validation, dissemination, and community building. Bibliometric analysis of journals transcends simplistic impact metrics to examine how they shape intellectual trajectories, enshrine theoretical frameworks, and establish disciplinary boundaries. The distribution of scientific output typically follows patterns of concentration, in which a small core of journals captures disproportionate attention, while an extensive periphery hosts specialized contributions with segmented audiences.

The invisible architecture of scientific fields is revealed by mapping journal networks through co-citation and bibliographic coupling analyses. These analyses identify clusters of publications that represent distinct epistemological subfields or methodological traditions, showing how specialized communication is organized within each discipline.

The relative position of a journal within these networks provides more nuanced insights into its intellectual influence than any one-dimensional metric, revealing its role as a bridge between communities, a bastion of orthodoxies, or a space for disruptive innovation.

The ecology of journals is undergoing profound transformations driven by digitization, open science, and new economies of academic attention.

The emergence of interdisciplinary mega-journals challenges established disciplinary taxonomies, while thematic fragmentation creates spaces for ultra-specialized publications. Simultaneously, the transition to open-access models is reconfiguring power relations among publishers, authors, institutions, and scientific societies, with as yet uncertain consequences for the quality, diversity, and sustainability of academic communication.

Health indicators for the journal ecosystem must consider dimensions beyond citation impact, including geographic and linguistic diversity, equity in editorial processes, and economic sustainability. Advanced bibliometric analysis allows for the diagnosis of distortions such as citation endogamy, thematic homogenization, and excessive concentration, informing policies for a more robust, diverse, and responsive publication ecosystem that meets the evolving needs of scientific communities.

9.3. Interpretation of distribution by country

The geographical distribution of scientific output reveals profound inequalities in research capacity at the global level, with a small group of countries concentrating the majority of high-impact publications and citations. These patterns of concentration reflect historical asymmetries in the distribution of research resources, the capitalization of scientific infrastructure, and the capacity to train advanced human capital. However, diachronic analysis shows gradual convergence processes driven by aggressive science policies in emerging economies and the growing internationalization of research collaboration.

Patterns of thematic specialization by country constitute a particularly revealing analytical dimension of comparative advantages in different national innovation systems. Some countries develop publication profiles that are highly specialized in specific fields where they have competitive advantages based on natural resources, established intellectual traditions, or specific industrial clusters. Others show more diversified profiles that reflect deliberate strategies for developing general scientific capabilities. Analysis of these specialized niches enables identification of excellence niches and opportunities for complementary international collaborations.

International collaboration networks represent the most dynamic dimension of contemporary knowledge geopolitics. Bibliometric analysis of international co-authorship reveals complex patterns of preferential association that typically reflect linguistic, historical, or geographical proximities, but also the emergence of new strategic alliances based on thematic complementarities or shared resources.

The position of countries within these global collaboration networks, as central hubs,

peripheral bridges, or isolated actors, significantly conditions their ability to access frontiers of knowledge and participate in transformative innovations.

9.4. Interpretation of gender in scientific authorship

A gender analysis of scientific authorship reveals deep patterns of differential participation in knowledge production. Gender distributions vary significantly across disciplines, reflecting particular histories of inclusion/exclusion, specific disciplinary cultures, and different paths to professionalization. In fields such as nursing or education, female authorship tends to be in the majority, while in engineering or physics, marked underrepresentation persists.

These disciplinary differences must be interpreted in light of historical, social, and institutional factors that have shaped the access and permanence of different genders within each field.

The authorship position is a crucial analytical dimension, as it reflects hierarchies of intellectual contribution and leadership in research projects. Consistent studies show that women are overrepresented in intermediate authorship positions and underrepresented as first authors or corresponding authors, particularly in high-impact international collaborations. These disparities point to possible barriers in the recognition of scientific leadership and in the assignment of central roles in research projects. These aspects deserve attention in gender equity policies in science.

The evolution of participation by gender over time shows complex trajectories with notable progress but also persistent stagnation. While some disciplines have experienced significant convergence, others continue to exhibit persistent gaps, particularly at the highest levels of productivity and impact. Diachronic analysis allows us to identify turning points linked to equity policies, cultural changes, or institutional interventions, providing valuable evidence for designing effective strategies to promote gender equality in the scientific system.

9.5. Interpretation of the documentary typology

The typological diversity of scientific publications reflects the plurality of discursive genres and communication forms within each discipline. Empirical research articles dominate in the natural and medical sciences, while books and book chapters remain central in the humanities and social sciences. These typological differences respond to epistemological traditions, knowledge validation practices, and academic reward structures specific to each field, and must be critically considered in any evaluative exercise.

Analysis of typological distribution over time reveals transformations in scientific communication practices.

The growing predominance of research articles in many disciplines reflects homogenization processes driven by evaluation systems based on journal metrics. However, alternative forms such as preprints, datasets, software, and other research products are resurging in response to open science movements and in recognition of the diversity of contributions to advancing knowledge.

The relative impact of different document types varies substantially across disciplines and subfields. While systematic review articles receive greater citation attention in some areas, methodological or theoretical contributions are more influential in others. Understanding these variations is essential for fair bibliometric evaluations that recognize the differential value of

different types of scientific contributions according to the validation criteria specific to each epistemic community.

9.6. Interpretation of keyword analysis

Keyword analysis provides a privileged window into the conceptual and cognitive structure of scientific fields. The frequencies of terms reveal the central concepts that define a research domain, while co-occurrences show the semantic connections that structure the conceptual space. This analysis allows us to identify consolidated thematic nuclei, interfaces between subfields, and conceptual gaps that represent opportunities for innovative research.

Terminological evolution captures processes of conceptual change, the emergence of new paradigms, and transformations in the theoretical frameworks of disciplines. Diachronic analysis of keywords enables us to track the migration of concepts across fields, the emergence of specialized vocabularies, and the disappearance of obsolete terms. These patterns of terminological change reflect deeper dynamics of reconfiguration of the cognitive landscapes of disciplines.

The analysis of terminological diversity offers insights into the degree of specialization or interdisciplinarity of a field of research. Highly specialized fields exhibit dense, shared technical vocabularies, while interdisciplinary areas exhibit greater terminological heterogeneity and conceptual borrowing from multiple domains. This analytical dimension is particularly valuable for studying processes of disciplinary fragmentation, the emergence of hybrid fields, and dynamics of scientific convergence.

9.7. Analysis of institutional collaboration networks

Institutional collaboration networks reveal the organizational architecture of the scientific system, showing how different types of institutions (universities, research centers, companies, hospitals) connect to produce knowledge. The analysis of these networks allows us to identify leading institutions that function as centers of collaboration, patterns of preferential association based on geographical proximity or thematic complementarity, and the emergence of strategic consortia that reconfigure the research landscape.

The evolution of collaboration networks shows trends toward increasing institutional interconnection, but with patterns marked by structural inequalities. Institutions in central countries typically occupy more central positions in global networks, while peripheral institutions face greater barriers to integrating into international collaborations. Diachronic analysis of these networks allows us to evaluate the impact of scientific policies designed to promote more inclusive and diverse collaborations.

Institutional diversity in collaborations is an indicator of the health of the research ecosystem. Collaborations involving multiple types of institutions (academia, industry, and government) typically generate knowledge with greater potential for social and economic impact. The analysis of this collaborative diversity informs policies aimed at strengthening links between different sectors of the innovation system.

9.8. Funding patterns in scientific production

Analysis of funding agencies and programs reveals the economic infrastructure that supports scientific research. The distribution of funding sources shows the centrality of certain agencies in supporting research, the thematic priorities that guide R&D investments, and the concentration or diversification of sources of support for science. This analysis provides valuable evidence for

evaluating the effectiveness of scientific funding policies.

The differential impact of different funding types is an analytical dimension of growing interest. Bibliometric studies show that competitively supported research typically achieves greater visibility and impact than research without explicit funding. In addition, the amount, duration, and type of funding (individual vs. collaborative, basic vs. applied) correlate with distinct patterns of productivity and impact, providing crucial information for optimizing research investment strategies.

Transparency in funding sources has increased significantly in recent decades, facilitating the analysis of the relationships between R&D investment and scientific results. This transparency allows connections to be drawn between funding priorities and emerging research topics, contributing to more robust accountability of science support systems and better alignment between scientific investment and social needs.

9.9. Combinations

The true power of contemporary bibliometric analysis emerges when multiple complementary dimensions are integrated to construct multidimensional analytical narratives. These combinations allow us to transcend the limitations of one-dimensional analyses, revealing complex interactions between different aspects of the scientific ecosystem. The systematic integration of dimensions such as gender, document type, collaboration patterns, and funding generates perspectives that would be impossible to obtain by examining each dimension separately, facilitating a more holistic and nuanced understanding of research dynamics.

The combination of gender analysis with institutional collaboration studies reveals crucial patterns of differential participation in scientific networks. For example, it is possible to examine whether female researchers participate in collaborative networks with the same structural characteristics as their male colleagues, or whether there are systematic differences in the type of institutions with which they collaborate. This integration enables identifying whether specific collaborative configurations, such as international multi-institutional consortia, exhibit higher or lower levels of gender equity, which is valuable for designing more inclusive scientific collaboration policies.

The intersection between typological analysis and temporal evolution reveals profound transformations in scientific communication practices. By combining these dimensions, we can track not only changes in production volumes but also transformations in the predominant discursive genres in different periods. This integration allows us to identify moments of paradigmatic change in which new forms of scientific communication emerge, or processes of homogenization in which certain types of documents displace others, with implications for the epistemological diversity of disciplines.

The triangulation of keyword analysis with funding data reveals the connections between funding priorities and the evolution of research agendas.

This combination allows us to examine how the introduction of new funding programs is reflected in the emergence of specialized terms in the literature, or how changes in funding agencies' priorities reorient the thematic focus of entire scientific communities. This analysis is particularly valuable for assessing the impact of science policies on the configuration of disciplinary cognitive landscapes.

The integration of geographical, thematic, and collaborative dimensions generates complex maps of the geopolitics of knowledge. By combining these dimensions, it is possible to identify not only which countries produce knowledge in which areas, but also how global collaboration networks are structured around different thematic specialties.

This integration reveals patterns of cognitive dependency, regional excellence niches, and asymmetries in countries' capacities to influence global research agendas.

Multivariate analysis of complementary dimensions faces significant methodological challenges, particularly in handling complex interactions and identifying causal relationships. However, approaches such as multiple correspondence analysis, structural equation modeling, and machine learning techniques allow these interrelationships to be systematically explored. The development of interactive panels that will enable different dimensions to be dynamically crossed represents a promising innovation for the visual exploration of these complexities.

The interpretation of integrated analyses requires particular sensitivity to disciplinary and contextual specificities. Patterns that appear universal may fade when examining individual disciplines, while relationships that seem marginal at the aggregate level may prove crucial in specific fields.

This contextual sensitivity is essential to avoid undue generalizations and to produce knowledge that respects the epistemological diversity of the scientific universe.

Combinations of dimensions open up possibilities for new synthetic metrics that capture multidimensional aspects of scientific production. Indicators of cognitive diversity, equity in collaborations, or alignment between funding and results emerge naturally from these integrations, offering more sophisticated tools for scientific evaluation.

Obtaining data for multidimensional bibliometric analysis is a highly contextual process that varies substantially depending on the specific research objectives and analytical questions posed. Each researcher must design a customized collection strategy that responds to their particular needs, typically starting from two primary sources: CSV files exported directly from bibliographic databases such as Scopus, Web of Science, or Dimensions, or raw results generated by specialized tools such as Publish or Perish, Bibliometrix, or VOSviewer. This initial choice profoundly shapes subsequent analytical processing, as each source has distinct formats, structures, and levels of complexity.

Once all the results are ready and you understand what each one means, the next step is to communicate them. The following section will address how to write a bibliometric report and prepare its presentation.

9.10. Visualization strategies for different bibliometric dimensions

The selection of appropriate visual representations transcends the merely aesthetic and becomes a fundamental methodological decision that directly affects the understanding and interpretation of bibliometric findings. Each type of analysis requires specific formats that highlight its characteristic patterns, facilitating the extraction of significant elements from complex datasets. The communicative effectiveness of a bibliometric study depends critically on this strategic choice, as the correspondence between the data structure and the visual format determines the recipient's ability to decode the information presented.

For analyzing linguistic distribution in scientific production, vertical or horizontal bar charts offer an optimal solution, enabling immediate comparisons of publication frequencies by language. This representation clearly visualizes the predominance of English as *a* scientific *lingua franca* over other languages, while allowing identification of temporal trends through historical series. The incorporation of color coding by language families or geographic regions adds analytical layers, transforming a simple quantitative distribution into a tool for exploring cultural hegemonies in global scientific communication.

The representation of geographical distribution finds its most intuitive and effective expression in choropleth maps, transforming tabular data into immediately understandable spatial narratives. This cartographic modality assigns color intensities according to the density of scientific production by country or region, revealing patterns of knowledge concentration and North-South asymmetries with immediate visual impact. Overlaying indicators of international collaboration through directed flows creates multidimensional visualizations that simultaneously capture individual productivity and cooperation networks, offering an integrated perspective of the global scientific landscape.

Specialized bibliometric tools such as Bibliometrix in R and PyBibX in Python automatically generate advanced visualizations that integrate multiple dimensions addressed in this chapter, combining them with bibliometric indicators such as the gross or annual citation rate.

The choice of which type of tool and representation to use is determined by the needs of what you want to communicate, always bearing in mind that you should use the most straightforward and most intuitive form of representation possible (plain text -> table -> graphic representation), without repeating information.

Recap

- Contextual dimensions complement the understanding of bibliometric results.
- The language of publication influences international visibility.
- Document type determines representativeness and weight.
- It is essential to standardize institutional affiliations.
- Gender analysis requires ethical and transparent methods.
- Gender inferences must be substantiated and documented.
- Country-specific analyses allow for the identification of regional inequalities.
- Linking funding to production helps evaluate impact.
- Industry-academia collaboration is measured in co-authorships and patents.
- Analysis by the journal shows editorial concentration.
- Crossing dimensions reveals structural inequalities.
- Missing data limits accuracy.
- Complementary sources (Crossref, OpenAlex) improve metadata.
- Geographic visualizations facilitate global comparisons.
- Analyzing languages in titles/abstracts measures accessibility.
- Database coverage influences biases.
- Normalization by field and year is essential for comparison.
- Uncertainty and data gaps must be reported.
- The confidentiality of subgroups must be respected.
- Incorporating additional variables enriches interpretation.

Self-assessment questions

1. Why is it useful to study the language of publication?

2. What biases arise from the exclusive use of Scopus or WoS?
3. How are affiliations normalized?
4. What ethical precautions are there in inferring gender?
5. How would you link funding data with topics?
6. What are the implications of document typology?
7. How can errors due to missing data be avoided?
8. Why use complementary sources?
9. What does crossing dimensions (country, gender) contribute?
10. Which visualizations are most effective?

BIBLIOGRAPHY

1. Gingras Y. Bibliometrics and Research Evaluation: Uses and Abuses. Cambridge (MA): MIT Press; 2016. ISBN: 9780262337663.

2. De Bellis N. Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Lanham (MD): Scarecrow Press; 2009. ISBN: 9780810867130.

3. Rousseau R, Egghe L, Guns R. Becoming Metric-Wise: A Bibliometric Guide for Researchers. Cambridge (MA): Chandos/Elsevier; 2018. ISBN: 9780081024744.

4. Waltman L. A Review of the Literature on Citation Impact Indicators. J Informetrics. 2016;10(2):365–91. DOI: 10.1016/j.joi.2016.02.007.

BIBLIOGRAPHIC REFERENCES

1. UNESCO. The race against time for smarter development | 2021 Science Report. https://www.unesco.org/reports/science/2021/en