Chapter 3 / Capítulo 3

Applied bibliometrics. From data to publication (English Edition)
ISBN: 978-9915-9680-7-0
DOI: 10.62486/978-9915-9680-7-0.ch03

Tools for Bibliometric Analysis/ Herramientas para el análisis bibliométrico

3.1. Introduction to bibliometric tools

Bibliometric software represents the methodological backbone that transforms raw data into actionable knowledge. Without these specialized tools, researchers would face the practical impossibility of manually analyzing the massive volumes of scientific publications that characterize the contemporary era. Automation allows thousands of references to be processed in minutes, identifying patterns of collaboration, thematic trends, and citation networks that would otherwise remain hidden in the noise of information. This computational capacity not only optimizes time resources but also fundamentally expands the frontiers of what can be analyzed, enabling research questions that were previously unattainable due to technical and operational limitations to be addressed.

The choice of the right software directly determines the validity and depth of bibliometric findings. Each tool incorporates particular methodological assumptions that condition the types of analysis possible and the interpretations derived. Platforms such as VOSviewer prioritize network visualization, while Bibliometrix emphasizes statistical indicators, and Python offers algorithmic flexibility. This specialization creates a complementary ecosystem where the strategic combination of tools overcomes individual limitations. Competence with multiple software programs thus becomes a fundamental skill for the contemporary bibliometrist seeking to produce robust, multidimensional scientific evaluations.

The evolution of bibliometric software reflects and simultaneously drives the theoretical development of the discipline. Current tools incorporate advances such as natural language processing for content analysis or machine learning algorithms for detecting emerging topics. This symbiosis between technological development and methodological sophistication has transformed bibliometrics from a basic descriptive exercise into a predictive analytical science. Modern researchers must understand both the bibliometric fundamentals and the technical capabilities of the software to design studies that leverage the full analytical potential of the contemporary digital ecosystem.

This book does not aim to provide an in-depth look at the tools used to create bibliometrics, but rather to equip the reader with the minimum knowledge needed to use them, along with guidance for those who wish to explore them in greater depth.

3.1.1. Classification of tools and selection criteria

The current bibliometric software ecosystem is highly diverse, responding to different methodological needs and levels of expertise. This variety can be organized into clearly defined categories according to technical complexity and specific analytical objectives. Understanding this fundamental taxonomy is the first step toward making an appropriate selection that optimizes available research resources.[1]

At the base of the tool pyramid is basic analysis software, designed for users who require quick calculations of fundamental indicators. These solutions prioritize accessibility over analytical depth, offering intuitive interfaces that minimize the learning curve. Their main advantage lies in their ability to deliver immediate results without specialized technical knowledge, though their methodological flexibility is limited for complex research.

Comprehensive analysis platforms represent the next level of technological sophistication, integrating multiple functionalities into unified environments. Often developed as complete suites, they allow complete bibliometric workflows to be executed without changing environments. Their modular architecture supports everything from fundamental descriptive analysis to advanced text mining and thematic pattern detection across large volumes of scientific literature.

Specialized visualization tools occupy a particular space within the ecosystem, focusing on the graphical representation of networks and knowledge structures. Their added value lies in specific algorithms for network layout and cartographic techniques that transform bibliometric data into interpretable maps. These solutions complement rather than replace the quantitative analysis performed on other platforms.

At the pinnacle of technical complexity are solutions that can be customized through programming, which give total control over every methodological aspect. This code-based approach sacrifices immediacy for absolute flexibility, allowing the implementation of novel techniques not available in commercial tools. It is the preferred option for cutting-edge bibliometric research that requires original methodological developments.

The selection between these categories should be guided by specific criteria that transcend personal preferences. Data volume is a primary consideration, as tools designed for meta-analysis of thousands of records differ significantly from those optimized for focused bibliographic studies. The software's scalability determines its suitability for projects of varying scope.

Available computational resources represent another decisive factor that is often underestimated in research planning. Some advanced visualization tools require substantial graphics processing capabilities, while web-based solutions externalize these technical requirements. A realistic assessment of available technological infrastructure prevents bottlenecks during the analytical phase.

Interoperability with specific data sources deserves special consideration, given the current fragmentation of the bibliographic ecosystem. Native compatibility with Scopus, WoS, Dimensions, or PubMed directly conditions the technical feasibility of many projects. Solutions that offer multiple, standardized connectors significantly reduce preprocessing and normalization efforts.

The learning curve for each tool must be weighed against the time available and the research team's expertise. While some modern platforms prioritize user experience through intuitive graphical interfaces, others require specialized knowledge that can slow down the initial stages of the project. Existing training and training capacity are critical variables in this equation.

Reproducibility and methodological transparency emerge as decisive criteria for high-impact research. Tools that allow the export of complete workflows and executable scripts facilitate peer review and the replicability of studies. This feature is particularly valuable in scientific evaluation contexts where auditability is a fundamental requirement.

3.2. Preparing the work environment for R

Some of the giants of bibliometric research are the R Bibliometrix(1) and Python PyBibx(2) libraries, respectively. Both languages require prior configuration and installation to be used. These will be addressed below.

3.2.1. Installation and configuration of R and RStudio

Preparing the R environment to run code in R begins with downloading the latest version of R from the Comprehensive R Archive Network (CRAN) (https://cran.r-project.org/), then selecting the installer for the operating system. After completing the basic installation, download RStudio Desktop (https://posit.co/products/open-source/rstudio), the integrated development environment that optimizes R workflows. The initial configuration includes defining the working directory, adjusting appearance preferences, and setting performance options based on the computer's capabilities. These are selected as they are displayed in the installer's options. This solid foundation ensures a stable environment for running bibliometric packages.

A fundamental concept to understand when working with these programming languages is that of a library or package, a collection of pre-written code that extends the language's basic capabilities. They function as a set of specialized tools, providing functions and methods for specific tasks without requiring developers to create them from scratch. For example, a data science library may contain algorithms for statistical analysis or graphics. This not only greatly speeds up the development process but also ensures reliability and efficiency by leveraging code that the community has already tested. In essence, they are fundamental components that allow complex applications to be built in a modular and efficient manner.

Package management in R is a crucial step for bibliometric analysis. It is recommended to start by installing essential packages, such as Bibliometrix, bibliometR, and the tidyverse, from the official CRAN repository. This process will be discussed in more detail later. Configuring personal libraries avoids version conflicts and facilitates project portability. This meticulous preparation prevents errors during subsequent analytical processes.

3.2.2. Specialized tools in R

Bibliometrix: Complete installation and configuration

Bibliometrix is installed in RStudio using the command install.packages("bibliometrix"). After installation, loading the library with the library(Bibliometrix) enables all its features for comprehensive bibliometric analysis.

Figure 3.1. Installation of R Bibliometrix

The Bibliometrix package offers three different interfaces to suit different user profiles.

The biblioshiny() function launches a web-based graphical interface ideal for beginners or quick exploratory analyses. For intermediate users, basic functions such as convert2df() and biblioAnalysis() provide direct control over the workflow. Advanced users can access all functions via direct programming in R, enabling sophisticated methodological customization. This flexibility makes Bibliometrix a complete solution for all types of bibliometric projects. In this text, the last two options will be used for greater control over the generated content.

3.2.3. Main features and fundamental analysis

Bibliometrix deploys a comprehensive analytical repertoire that begins with fundamental descriptive analysis of scientific output. The biblioAnalysis function automatically generates fundamental indicators, including temporal growth, most productive authors, and leading journals. The identification of collaborations using networkPlot reveals co-authorship patterns at the individual, institutional, and national levels. These descriptive functionalities provide the necessary contextual basis before undertaking more specialized analyses such as conceptual mapping or citation studies.

Content analysis and thematic trends are among Bibliometrix's most powerful capabilities. The conceptual structure function uses co-word analysis to identify thematic clusters and their evolution over time. Co-citation network mapping using cocMatrix reveals the intellectual structure of the field of study. The thematicMap function visualizes topics according to their degree of development and relevance, distinguishing between emerging, basic, niche, and driving issues. These tools facilitate the identification of knowledge frontiers and research opportunities. How to use and interpret them will be discussed in more detail in later chapters.

The R ecosystem for bibliometrics is enriched with specialized libraries that complement Bibliometrix. One of these is bibliometR, which offers additional functionalities for citation network analysis and community detection. RefManageR facilitates advanced management of bibliographic references in different citation styles. There are others, but currently the most widely used is RBibliometric, which will be the focus of this book.

3.3. Working with Google Colab for PyBibX

Google Colab is the ideal platform for implementing PyBibX, eliminating the need for local configuration. Free access via a Google account immediately provides a preconfigured Python environment with the main scientific libraries. Connection to enhanced computational resources, including GPUs and TPUs, significantly speeds up the processing of large bibliometric datasets. This cloud infrastructure ensures consistency among collaborators and simplifies the reproducibility of analyses.

Since Google Colab runs in the browser, no prior equipment configuration is necessary; only internet access from the device's browser is required. It will only be essential to install the required libraries, as in the case of RBibliometric. This is done with the *pip install pybibx command*.

The official PyBibX documentation (https://pypi.org/project/pybibx/) provides a list of codes for the types of files to be imported for analysis, so it is not necessary to memorize each code or write them manually. Open the link according to the file and start working.

There are other alternatives to Colab, such as Anaconda. Still, in this text, Colab will be used because its session management optimizes work with PyBibX in extensive bibliometric projects, the connection to runtime with high RAM prevents failures during the processing of large document corpora, and the scheduling of periodic executions using cron jobs allows analyses to be kept up to date without constant manual intervention. Automatic export of results to Google Drive ensures the persistence of findings beyond the active session on the platform.

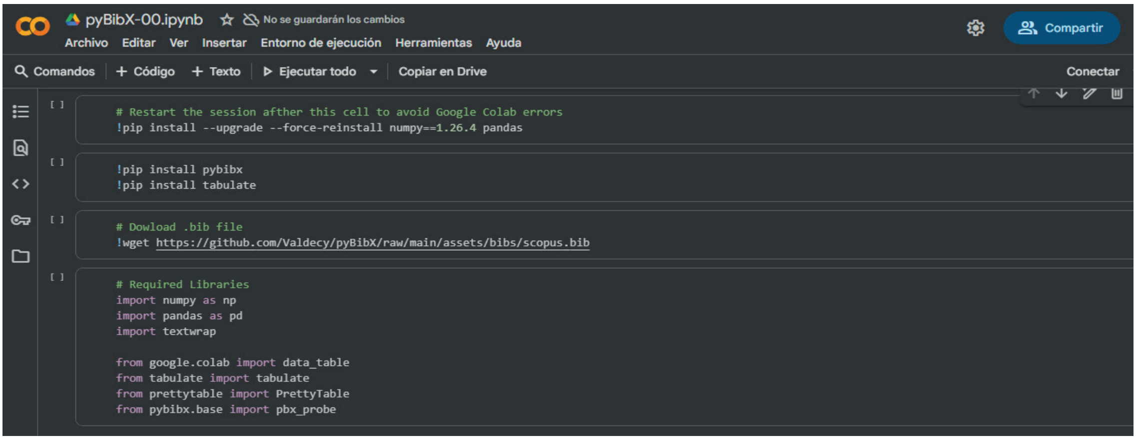


Figure 3.2. PyBibx project in Google Colab

In both R Bibliometrix and Pybibx, after importing the libraries, it is necessary to load the data file (s) for the study. Each of these libraries has its own particular way of importing and subsequently analyzing them.

R Bibliometrix

In R (Bibliometrix), the convert2df() function reads an exported file from databases such as Scopus or WoS and converts it into a data frame that the library can work with.

```
# Load the library
library(bibliometrix)

# Import and convert the file downloaded from Scopus/WoS into a dataset called my_dataset
my_dataset <- convert2df(“my_scopus_file.bib”, dbsource = “scopus”, format = “bibtex”)
```

Note: anything after # is a comment; even if you paste it into RStudio or Colab, it will not be executed as code.

Pybibx

In Python (Pybibx) with Google Colab, you must first upload the file to the environment and then use the load function of the dataset module to load it into a Collection object, which is the central data structure. This may require other library imports. The official Pybibx website provides a list of code examples for each type of data file, for example, if they were exported as .bib from Scopus, or txt from PubMed, among others, depending on their source.

```
# Install and import the library (in Google Colab)
!pip install pybibx
from pybibx import dataset

# Upload the .bib or .csv file to Colab
from Google. colab import files
```

```
uploaded = files.upload()

# Load the dataset into a Collection called my_dataset
my_dataset = dataset.load(list(uploaded.keys())[0])
```

You don't need to know how to program to use this book; the code shown can be copied as is and will work.

3.4. Publish or Perish: Analysis with Google Scholar

3.4.1. Installation, configuration, and search types

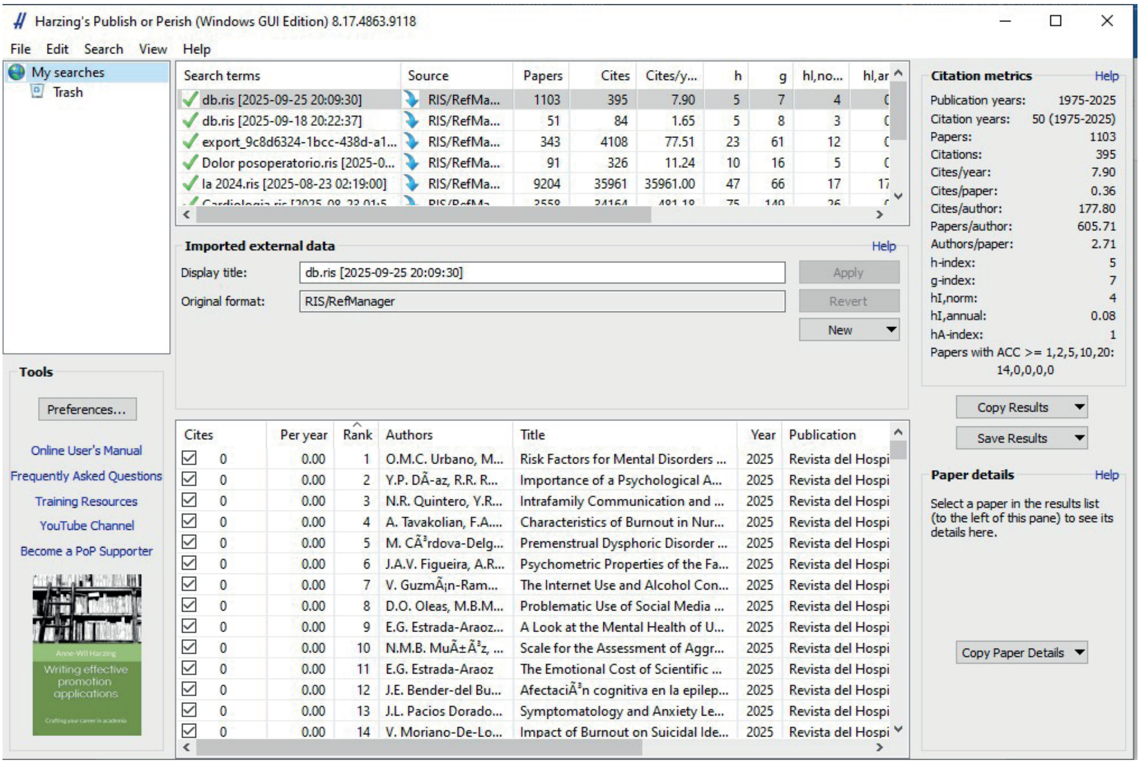


Figure 3.3. Harizing's Publish or Perish home screen

The installation of Publish or Perish begins with a free download from the official Harzing.com website (https://harzing.com/), selecting the version compatible with the user's operating system. The installation process is simple and does not require advanced technical configuration, though it is essential to allow network connections so the software can access Google Scholar data. After installation, the main interface displays clearly organized options that let you start performing bibliometric searches immediately after initial configuration.

Optimal software configuration involves customizing preferences to specific research needs. Among the most relevant settings is the selection of the primary search engine, where Google Scholar is the most commonly used default option. It is crucial to configure the wait time between queries to avoid being blocked by Google, setting intervals that simulate human behavior. Customizing export formats and result limits completes the preparation for efficient and sustainable use of the tool.

The search types available in Publish or Perish are tailored to different bibliometric objectives. The author search allows you to locate and analyze the impact of specific researchers using different variations of their names. The article search makes it easy to track citations of individual publications by combining title, author, and year. The general search by research terms provides a broad overview of the literature on particular topics, while the journal search evaluates the impact of specific periodicals.

Each search type requires specific strategies to optimize data retrieval. For authors, it is essential to try different combinations of names and initials to counteract Google Scholar's limitations in disambiguation. In thematic searches, the strategic combination of Boolean operators and exact phrases significantly improves the accuracy of results. Setting filters by publication year, language, and document type refines result sets before analysis, saving time in subsequent processing of retrieved information.

The software offers advanced search capabilities that overcome the limitations of the Google Scholar web interface. The ability to search in multiple languages extends the scope of bibliographic retrieval beyond Anglo-centric literature. The search function by institutional affiliation, although limited by Google Scholar's coverage, provides valuable insights into the productivity of different research centers. These features make Publish or Perish an indispensable tool for bibliometric analysis, complementary to commercial databases.

The effectiveness of searches depends largely on understanding Google Scholar's inherent limitations. Uneven coverage across disciplines, the inclusion of non-peer-reviewed documents, and duplicate records require careful interpretation of results. The software partially mitigates these problems through cleaning and deduplication algorithms, but the researcher's critical judgment remains essential to validate the quality of the data obtained before formal bibliometric analysis.

Publish or Perish significantly expands its capabilities by integrating direct connections to Scopus, Web of Science, PubMed, and other databases using valid institutional credentials. This functionality turns the tool into a bridge between Google Scholar metrics and the databases above, enabling unique comparative analyses. Users can authenticate themselves through institutional subscriptions to access standardized data of higher bibliographic quality. This integration allows for cross-referencing across sources, identifying discrepancies and complementarities in citation coverage for a more comprehensive assessment of research impact.

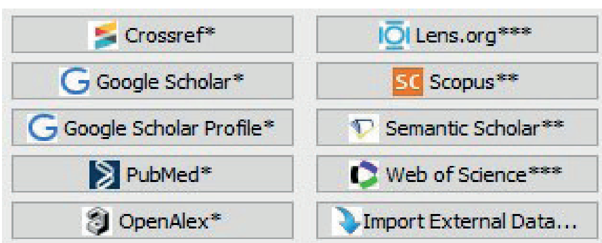


Figure 3.4. Publish or Perish data import block

The ability to import data from multiple sources is a distinctive advantage of Publish or Perish over other bibliometric tools. The software accepts standard formats such as RIS, BibTeX, and CSV from reference managers or exports from other platforms. This interoperability facilitates the consolidation of scattered bibliographic information into a single metric analysis environment. Researchers can combine datasets from different sources to create customized collections that fully reflect their scientific output, thereby overcoming the coverage limitations of individual sources.

The import process allows basic metadata to be enriched with impact indicators calculated by Publish or Perish. When loading references from Zotero, Mendeley, or Scopus exports, the software automatically searches for corresponding citations in Google Scholar and other available sources. This functionality is particularly valuable for updating existing bibliographic collections with recent citation data without manual searches. The result is an integrated dataset that preserves the original metadata while adding the most up-to-date impact metrics.

3.4.2. Data extraction and metrics provided

The Publish or Perish system performs comprehensive searches across Google Scholar, Scopus, and Web of Science, capturing essential metadata, including titles, authors, affiliations, publication years, and sources. The tool automatically processes citation networks to calculate advanced indicators, identifying both citations received and temporal impact trends. This comprehensive process ensures that complete datasets are obtained for rigorous bibliometric analysis.

The primary metrics automatically calculated include the h-index, g-index, and individual h-index, providing different perspectives on research impact. The software also generates the total number of citations, citations per year, citations per article, and average citations per work. These indicators are complemented by temporal statistics that reveal patterns of productivity and impact evolution throughout the research career. Each metric is presented with its detailed calculation to facilitate contextual interpretation.

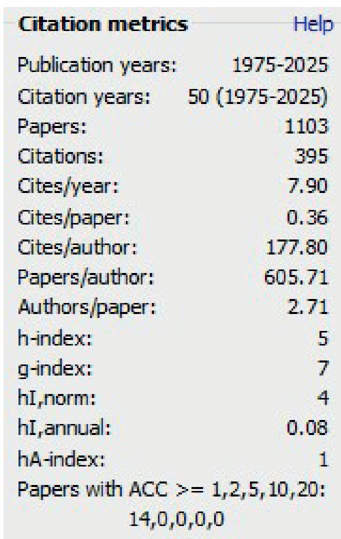


Figure 3.5. Publish or Perish analysis results block

The tool offers sophisticated comparative analyses that allow researchers, institutions, or journals to be evaluated against disciplinary benchmarks. The system calculates impact percentiles, relative positions, and comparative trends using standardized algorithms. This functionality is particularly valuable for bibliometric studies that require contextualizing performance within specific fields of knowledge. The results include clear visualizations that facilitate the identification of strengths and areas for improvement in the analyzed research profiles.

The extracted data includes qualitative information essential to correctly interpreting quantitative metrics. The software captures full titles, abstracts when available, and the most relevant cited references, allowing for an understanding of the reasons behind the measured impact, differentiating between citations of recognition, methodological citations, and academic controversies. The integration of qualitative and quantitative dimensions substantially enriches the resulting bibliometric analysis.

3.4.3. Export and integration with other tools

Publish or Perish offers advanced export capabilities that facilitate interoperability with the bibliometric ecosystem. Users can export complete results in CSV, RIS, or BibTeX formats, preserving all metadata and calculated metrics. This flexibility allows data to be integrated with specialized tools such as VOSviewer for network visualization, Bibliometrix for advanced statistical analysis, or CitNetExplorer for studying citation networks. The export maintains the relational structure of the data, ensuring that connections between authors, articles, and citations remain intact during transfer.

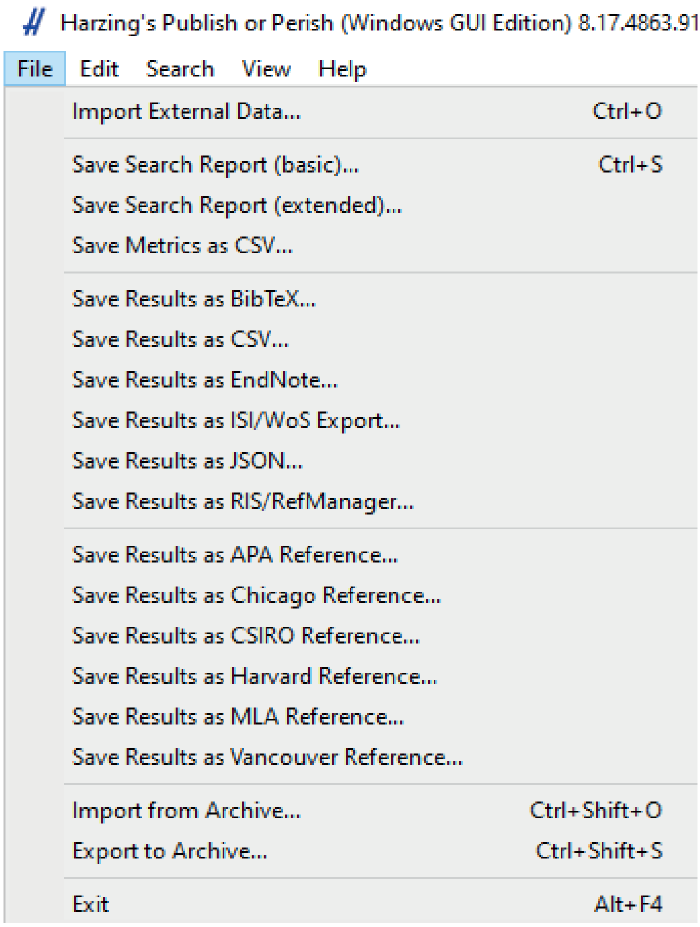


Figure 3.6. Publish or Perish export menu

Integration with bibliographic managers is a particularly valuable feature for researchers. Data exported in RIS format can be imported directly into Zotero, Mendeley, or EndNote, enriching personal libraries with calculated impact metrics. This feature eliminates the need for manual updating processes and ensures that bibliographic collections maintain up-to-date metric information. Bidirectional synchronization allows simultaneous work on reference management and bibliometric analysis within cohesive workflows.

For advanced quantitative analysis, exporting in CSV format provides immediate compatibility with statistical software such as R, Python, or SPSS. The resulting tables include standardized columns for each bibliometric variable, facilitating processing using custom scripts. Researchers can combine this data with additional information from other sources to create enriched datasets for multivariate models. This capability transforms Publish or Perish into an initial extraction tool within complex analytical pipelines.

Export options include customizable settings to suit specific research requirements. Users can select subsets of metrics, filter by time ranges, or choose particular encoding formats. This granularity ensures that data transferred to other tools contains exactly the information needed for each type of subsequent analysis. The combination of flexibility and precision in exporting positions in Publish or Perish is a fundamental component of integrated bibliometric workflows.

3.4.4. Limitations and best practices for use

Publish or Perish has significant limitations that users must recognize to interpret the results correctly. Its primary reliance on Google Scholar introduces inherent biases, including uneven coverage across disciplines, the inclusion of non-peer-reviewed documents, and record duplication. The tool lacks robust mechanisms for automatic author disambiguation, which can distort individual profiles when multiple researchers share the same name. Furthermore, the calculated metrics primarily reflect quantitative impact, without accounting for qualitative dimensions such as journal prestige or the type of scientific contribution.

Essential best practices include systematic cross-checking of results with complementary databases such as Scopus or Web of Science. Users should implement conservative search strategies, using multiple variations of author names and manually validating the most relevant profiles. It is crucial to contextualize metrics within specific disciplinary norms, recognizing that identical h-index values have different meanings in theoretical physics and philosophy. Methodological transparency requires documenting all search and data cleaning decisions in final reports.

Responsible interpretation of metrics requires understanding their methodological foundations and technical limitations. Users should complement quantitative metrics with qualitative assessment, reviewing the most-cited works to determine their actual contribution to the field. This balanced approach prevents erroneous conclusions drawn solely from numerical indicators without interpretive context.

The ethical management of the data obtained involves respecting the use limits established by the sources and avoiding overloading their servers. Massive searches should be scheduled with reasonable intervals between queries, and results should be stored in accordance with standard security protocols. Institutional users should establish clear guidelines for the appropriate use of metrics in academic evaluations, aligning with the principles of the DORA Declaration to avoid reductionism in research assessment.

3.5. Advanced visualization tools

The visualization of bibliometric data is a crucial stage in interpreting and communicating results, transforming complex datasets into intuitive graphical representations that reveal underlying patterns, relationships, and trends in the scientific literature. These specialized tools allow researchers and evaluators to identify collaboration networks, thematic structures, and knowledge dynamics that would remain hidden in conventional tables and lists, facilitating both specialized analysis and the effective dissemination of bibliometric findings to multidisciplinary audiences.

3.5.1. VOSviewer: installation and creation of scientific maps

The installation of VOSviewer begins with a free download from the official website of the Center for Science and Technology Studies at Leiden University, available for Windows, Mac OS, and Linux (https://www.VOSviewer.com/download). The installation process is simple and does not require advanced technical configuration. However, it is advisable to verify that the system has the latest version of Java to ensure optimal performance of all features. Once installed, the work environment is organized into logical modules that guide the user through the entire process of creating scientific maps.

Map creation in VOSviewer begins with the import of data from bibliographic sources such as Scopus, Web of Science, or PubMed, which have been previously processed into compatible formats. The next chapter will describe the process of obtaining this data. The software offers three main types of analysis: co-authorship networks to visualize scientific collaborations, co-citation networks to reveal the intellectual structure of a field, and term co-occurrence maps to identify research topics. Each type of analysis uses specific layout and clustering algorithms that optimize the visual representation of bibliometric relationships.

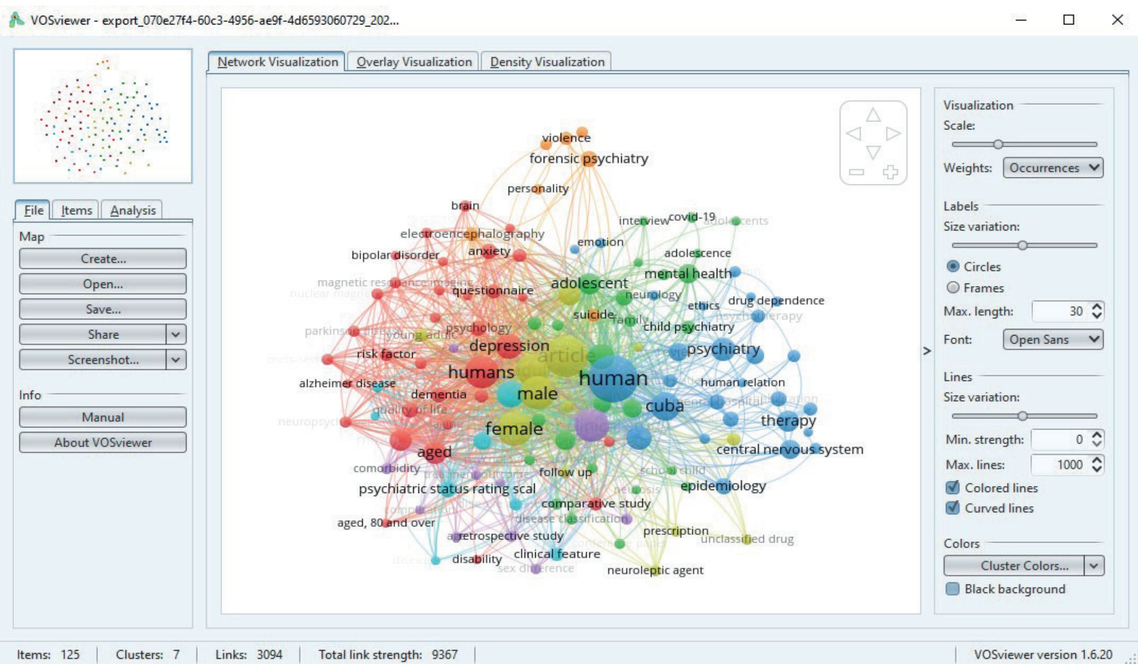


Figure 3.7. Main view of VOSviewer, with a project displayed

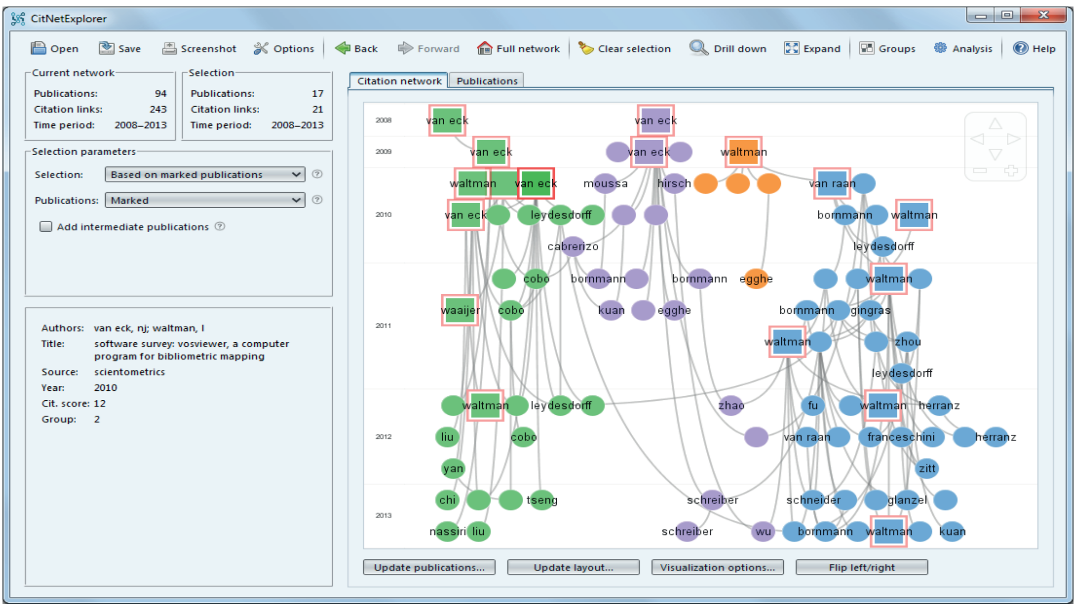
The mapping process involves several configuration stages in which the user defines key parameters, such as the minimum occurrence threshold, normalization method, and clustering algorithm. VOSviewer applies advanced dimensionality reduction techniques using the VOS (Visualization of Similarities) algorithm, which positions elements in a two-dimensional space while preserving their similarity relationships. The tool automatically generates color-coded clusters representing thematic communities or collaboration groups, facilitating the visual identification of structural patterns.

Map customization allows adjustment of multiple visual aspects, such as node sizes, link thicknesses, text fonts, and color schemes. Nodes can be sized according to different metrics, such as frequency of occurrence, number of citations, or centrality in the network, while link thickness reflects the strength of the relationships between elements. These customization options allow the visualization to be adapted to the specific needs of each analysis and target audience, improving the communicative clarity of the results.

Interpreting the generated maps requires understanding both the relative position of the elements and their grouping into thematic clusters. The distance between nodes indicates their degree of thematic or collaborative relationship, while the spatial distribution reveals the overall structure of the field of study. The maps allow for the identification of emerging themes, central authors, leading institutions, and patterns of international collaboration, providing valuable insights for scientific evaluation and future research planning.

VOSviewer includes advanced features such as temporal analysis using overlay maps, export to vector formats for academic publications, and zoom and filtering tools to explore specific areas of interest. Integration with other bibliometric software using standard exchange formats completes a robust visualization ecosystem that supports bibliometric research from initial exploratory analysis to the final presentation of complex results.

3.5.2. CitNetExplorer: citation network analysis



Source: CitNetExplorer official website (https://www.citnetexplorer.nl).
Figure 3.8. Main view of CitNetExplorer, with project graph

CitNetExplorer specializes in the analysis and visualization of citation networks, allowing users to explore the structure and evolution of scientific fields through the relationships between publications. Developed by Leiden University, this software enables users to identify key publications, trace the evolution of research lines, and analyze citation patterns over time. Its unique approach facilitates the study of intellectual heritage and knowledge trajectories within specific scientific disciplines.

The software is easy to install: download it from the official website and decompress the ZIP file into the desired directory (https://www.citnetexplorer.nl/download). CitNetExplorer does not require conventional installation; it runs directly from the JAR file with Java Runtime Environment 8 or higher. This portability allows the tool to be used across different systems without complex configuration, though it is recommended to verify execution permissions on Unix-based systems to ensure proper operation.

The analysis process begins with importing data from Web of Science or Scopus using specific export formats that preserve the cited references. The tool automatically constructs the citation network, where nodes represent publications and links represent citation relationships. Visualization algorithms organize publications chronologically, clearly showing the field's evolution and enabling the identification of seminal works that serve as connecting points between different lines of research.

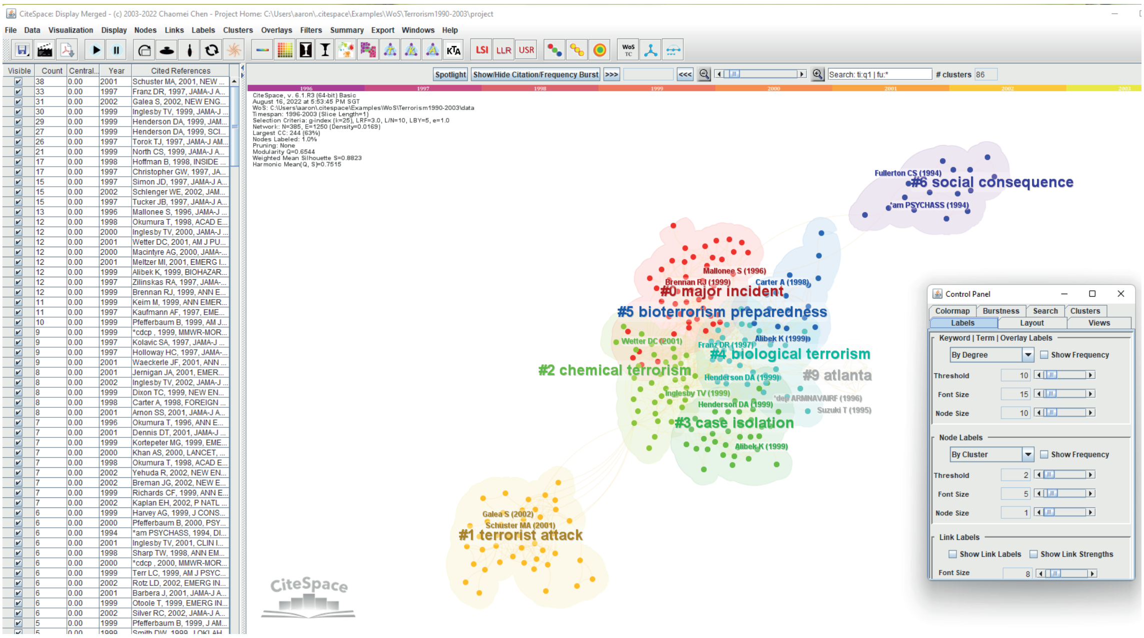
Exploration features include zoom and time-filtering tools that let users focus on specific periods. Users can select key publications to visualize their local citation network and identify predecessor and successor works. Citation path analysis reveals how scientific ideas are transmitted and transformed through different publications, providing insights into the dynamics of knowledge dissemination within the academic community.

The identification of thematic clusters is performed using algorithms that detect highly connected communities of publications. These groups represent specific subfields or lines of research, whose evolution can be tracked over time. CitNetExplorer allows you to analyze how different intellectual traditions emerge, converge, or diverge, offering unique insights into the structuring of scientific knowledge that complement analyses from other bibliometric visualization tools.

Export options include detailed reports with network metrics, high-resolution images for publications, and processed data for complementary analysis. The tool allows users to save complete work sessions, facilitating continuity in extensive research. This ability to preserve the analysis state is particularly valuable when working with large volumes of data or developing complex longitudinal studies of scientific evolution.

3.5.3. CiteSpace: analysis of temporal patterns

CiteSpace stands out as a tool that specializes in analyzing temporal patterns and detecting emerging trends in the scientific literature. Developed by Dr. Chaomei Chen, this software allows users to map the evolution of scientific domains using advanced data mining techniques and dynamic visualization. Its ability to identify citation bursts and turning points in scientific development makes it an invaluable tool for technology foresight studies and diachronic research analysis.



Source: CiteSpace official website (https://www.java.com/en/download).
Figure 3.9. Main view of CiteSpace, with project graphically represented

Installation requires the Java Runtime Environment (https://www.java.com/en/download) to be installed beforehand and is done using an executable JAR file available on the official website (https://citespace.podia.com/). The process includes configuring memory parameters to optimize performance with large volumes of data. CiteSpace organizes its interface into modular panels that sequentially manage data import, analytical processing, and result visualization, maintaining a structured workflow. However, it has a considerable learning curve for novice users.

Temporal analysis is based on the construction of citation networks segmented by user-defined periods. The software applies community-detection algorithms to identify thematic clusters and calculates centrality metrics to locate bridge publications across different areas of knowledge. The superimposition of consecutive networks allows the visualization of the emergence, convergence, or disappearance of lines of research over time, revealing the dynamics of scientific change.

The citation burst detection function identifies publications that experience sudden increases in citation frequency, signaling potentially revolutionary contributions or emerging research topics. These bursts are visualized by color-d rings on the network nodes, where the intensity and timing of the color indicate the magnitude and duration of the high citation period. This feature allows you to quickly identify works that have marked turning points in their field.

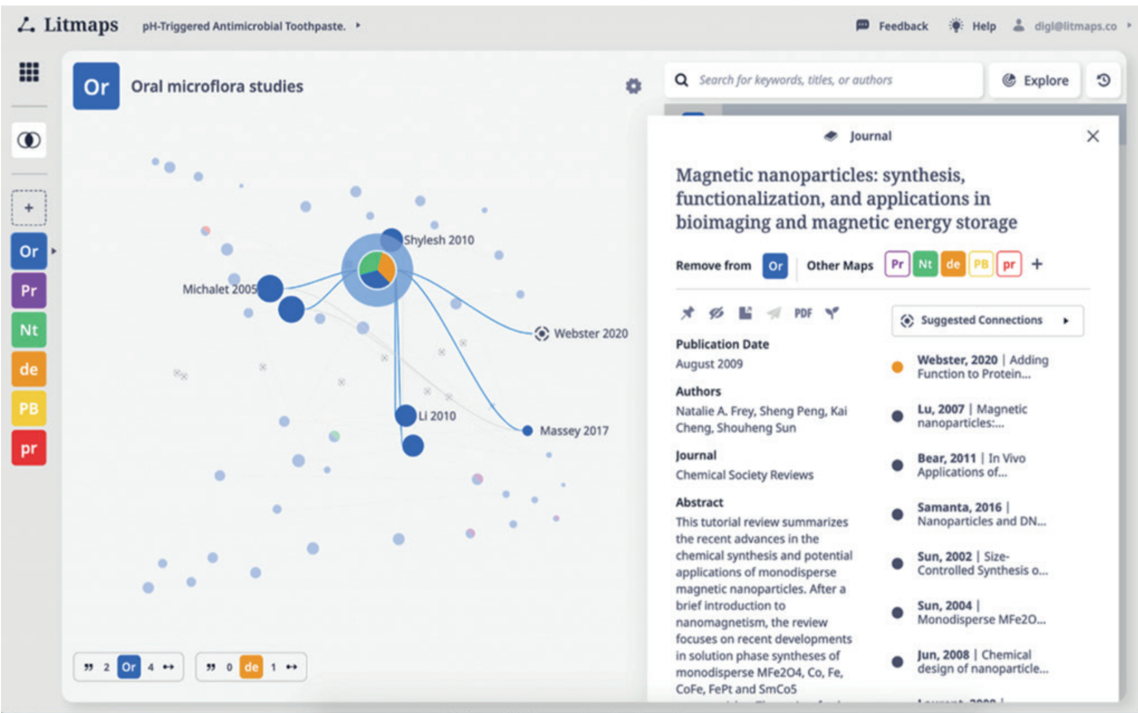
Scientific landscape visualizations use topographic models in which elevations represent connection density, and valleys indicate thematic discontinuities. CiteSpace generates evolutionary maps showing the drift of thematic clusters over time, complemented by trend lines that project future developments. These representations facilitate the identification of research opportunities and areas with growth and innovation potential.

The export of results includes detailed metric reports, animations of temporal evolution, and processed data for secondary analysis. CiteSpace generates tables of thematic clusters, including their fundamental publications, cohesion metrics, and representative labels, extracted using natural language processing algorithms. This comprehensive documentation supports the interpretation of detected patterns and facilitates the communication of complex findings on temporal dynamics in science.

3.5.4. Online tools (Litmaps, ResearchRabbit)

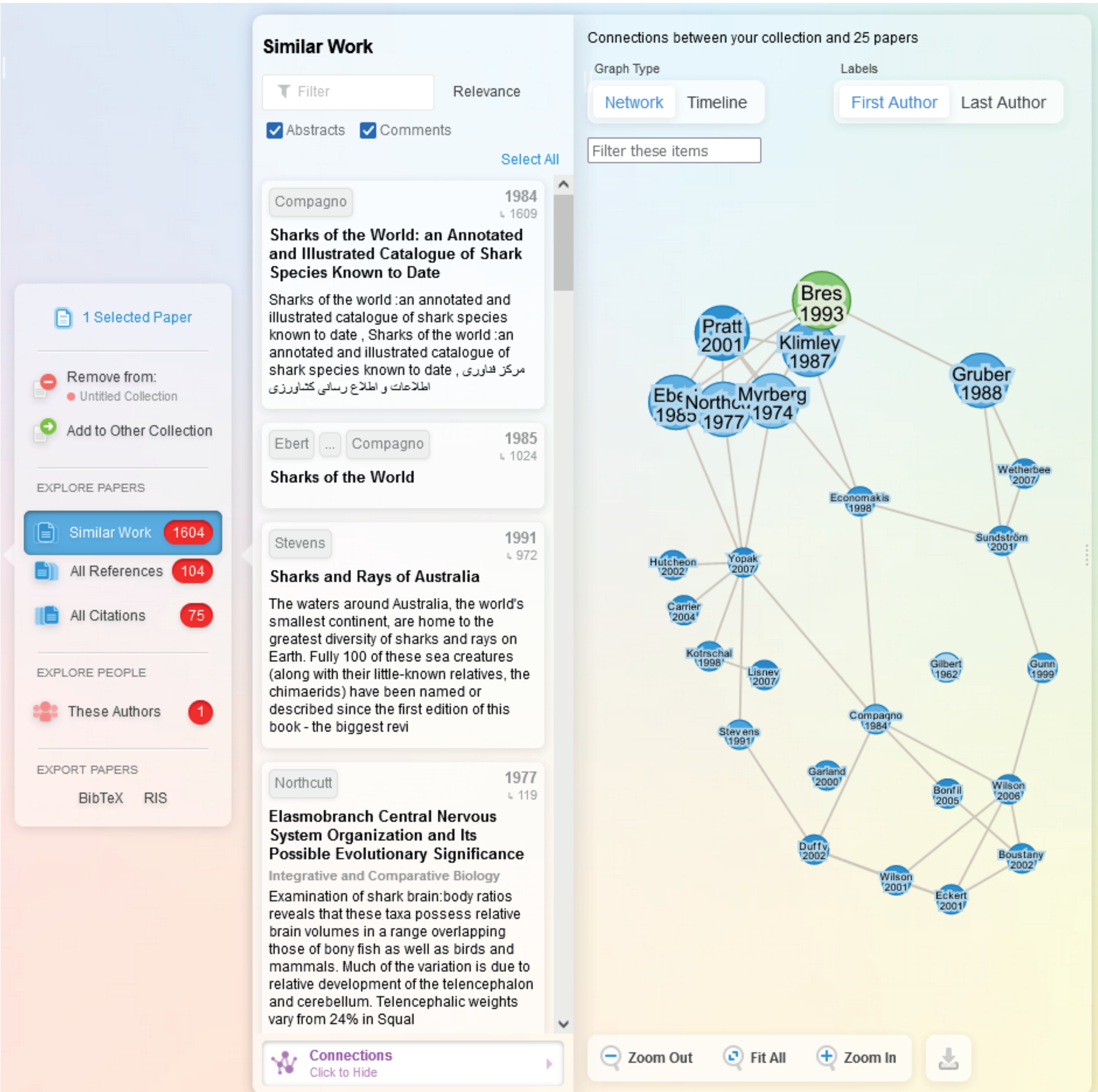
Online tools represent the latest evolution in bibliometric visualization, offering immediate accessibility and collaborative cloud-based functionalities. Litmaps and ResearchRabbit are emerging as innovative platforms that transform literature exploration through intuitive interfaces and advanced recommendation algorithms. These web solutions allow users to discover hidden connections between publications, track scientific developments in real time, and collaborate in geographically distributed research teams, marking a significant transition toward collaborative, real-time bibliometrics.

Litmaps (https://www.litmaps.com/) stands out for its ability to generate interactive citation maps that dynamically reveal relationships between publications. The platform allows users to visualize both the references cited in a seminal article and the subsequent citations of that article, creating knowledge networks that expand organically. Its "Seed Maps" algorithm automatically identifies the most relevant publications on a topic, while the alerts feature notifies users of new related research. This approach facilitates serendipitous literary exploration and the discovery of unexpected interdisciplinary connections.



Source: Litmaps official website (https://www.litmaps.com/).
Figure 3.10. Main view of Litmaps, with project graphically represented

ResearchRabbit (https://researchrabbitapp.com/) takes a different approach, functioning as "Spotify for academic research" through personalized recommendations based on reading preferences. The tool allows users to create collections of articles and receive increasingly refined suggestions through machine learning algorithms. Its timeline interface visualizes the historical evolution of research lines, while collaboration features enable users to share and comment on collections with research teams. This user-centered approach transforms literature exploration into a personalized and cumulative experience.



Source: ResearchRabbit official website (https://researchrabbitapp.com/).
Figure 3.11. Main view of ResearchRabbit, with project graphically represented

Implementing these tools requires only a modern web browser and free registration, eliminating barriers related to installation and operating system compatibility. Both platforms synchronize data in the cloud, allowing access from multiple devices and automatic recovery of work sessions. Integration with reference managers such as Zotero and Mendeley facilitates the export of relevant discoveries, while connections to major databases ensure comprehensive coverage of up-to-date literature.

Practical limitations include dependence on a stable internet connection and considerations regarding research data privacy. The closed nature of their algorithms can make it challenging to validate the methodology behind the recommendations, and free versions typically impose limits on the volume of analysis. Nevertheless, their immediate usability and ability to reduce the time spent exploring the literature make them valuable additions to the modern bibliometric toolkit, particularly in the early stages of research and for researchers in training.

3.6. Selection of tools

The appropriate selection of bibliometric tools requires a systematic evaluation that considers the specific research objectives, available resources, and the technical characteristics of each software tool. This chapter provides a comprehensive comparative framework to guide the selection of tools across different analysis scenarios, highlighting both individual capabilities and potential synergies between complementary platforms. The optimal strategy often combines multiple tools in integrated workflows that leverage each tool's distinctive strengths.

For mapping scientific collaboration networks, VOSviewer is the preferred choice thanks to its layout algorithms optimized for visualizing relationships between authors and institutions. When the analysis requires identification of communities and structural patterns in large networks, the combination of Bibliometrix for statistical processing and VOSviewer for visualization produces particularly robust results. In projects that prioritize communicative clarity over analytical detail, online tools such as ResearchRabbit offer immediately understandable visualizations for non-specialist audiences.

CiteSpace is the most powerful tool for analyzing citation networks and temporal evolution, especially for detecting turning points and emerging trends in rapidly evolving scientific domains. When the main objective is to track the historical development of specific ideas, CitNetExplorer provides superior diachronic capabilities for visualizing knowledge trajectories. For technology foresight studies that require identifying areas of research opportunity, combining CiteSpace with Litmaps' burst analysis offers valuable complementary insights.

In contexts of individual or institutional scientific evaluation, Publish or Perish allows for quick calculations of fundamental indicators from Google Scholar. At the same time, Bibliometrix offers a more comprehensive analysis by integrating multiple data sources. For evaluation committees that require standardized, comparable metrics, the combination of both tools provides immediacy and analytical depth. In institutional environments with limited technical resources, online tools have the advantage of requiring minimal infrastructure and prior training.

Effective integration of bibliometric tools follows the principle of modularity, in which each tool contributes its specific strengths to a coherent workflow. A common strategy combines data extraction with Publish or Perish, statistical processing with Bibliometrix or PyBibx, and advanced visualization with VOSviewer or CiteSpace. This pipeline approach maximizes individual capabilities while mitigating the particular limitations of each tool, producing more comprehensive analyses than are possible with any single platform.

Technical interoperability is facilitated by standardized exchange formats such as RIS, BibTeX, and CSV, which enable data transfer between tools with minimal loss of information. Custom scripts in R or Python can automate these conversions, which is especially useful when integrating tools based on different technological ecosystems. Meticulous documentation of the parameters used at each stage ensures the reproducibility of the integrated analysis and facilitates the identification of discrepancies between results obtained with different tools.

Integration strategies must consider both technical aspects and available human resources. Teams with programming expertise can implement highly customized workflows that combine Bibliometrix with complementary analyses in R or Python. For groups with limited technical capabilities, integrating tools with graphical interfaces, such as VOSviewer and Publish or Perish, offers an effective balance between analytical capabilities and practical usability. Cross-training in complementary tools is a strategic investment that significantly expands the research team's analytical capabilities.

Recap

- Bibliometric tools enable the extraction, processing, visualization, and analysis of scientific data from specialized databases.
- Their systematic use improves the accuracy and reproducibility of bibliometric studies.
- There are three main types of tools:
  - Data retrieval and extraction.
  - Statistical analysis and processing.
  - Scientific information visualization and mapping.
- The most widely used bibliographic databases are Web of Science (WoS), Scopus, Dimensions, PubMed, Lens, and Google Scholar.
- Web of Science offers classic indicators such as the Journal Impact Factor and is essential for longitudinal studies.
- Scopus, from Elsevier, includes the CiteScore indicator and the SCImago Journal Rank (SJR) system.
- Google Scholar is a valuable open-source resource for broad coverage analysis, though with less quality control.
- Notable analysis tools include VOSviewer, Bibliometrix (R), CiteSpace, SciMAT, and Gephi.
- VOSviewer is widely used to construct maps of word co-occurrence, co-authorship, and co-citation.
- Bibliometrix, an R package, enables advanced statistical analysis and exports to Biblioshiny, its interactive web interface.
- CiteSpace focuses on detecting trends and thematic clusters through temporal citation networks.
- SciMAT is used for evolutionary analysis of scientific production and for visualizing the dynamics of topics over time.
- Gephi allows complex collaboration and co-citation networks to be represented with a high degree of visual customization.
- Statistical and text mining tools (Excel, SPSS, R, Python) are also used to process exported data.
- The combination of several tools enhances the validity of the analysis and enables complementary perspectives.
- The selection of the tool depends on the study's objective, sample size, and desired level of detail.
- It is essential to maintain transparency and traceability throughout the data

extraction and cleaning processes.

- Bibliometric visualization facilitates the interpretation of results and scientific communication.
- Tools must be used with consistent ethical and methodological criteria, avoiding manipulation of indicators.
- Mastery of these tools is an essential skill for researchers, analysts, and science managers.

Self-assessment questions

1. What are the three main categories of bibliometric tools?
2. What functions do databases such as Web of Science and Scopus perform?
3. What indicator mainly characterizes Web of Science?
4. What metrics system does Scopus offer in addition to CiteScore?
5. What are the advantages and limitations of Google Scholar for bibliometric analysis?
6. What is VOSviewer used for in scientific network studies?
7. What features differentiate Bibliometrix from other tools?
8. What does CiteSpace contribute to the study of thematic trends?
9. How important is visualization in bibliometric analysis?
10. Why is it essential to document data extraction and analysis processes transparently?

BIBLIOGRAPHY

1. Moed HF. Citation Analysis in Research Evaluation. Dordrecht: Springer; 2005. DOI: 10.1007/1-4020-3714-7

2. Thelwall M. Web Indicators for Research Evaluation: A Practical Guide. San Rafael (CA): Morgan & Claypool; 2016. DOI: 10.2200/S00733ED1V01Y201602ICR048

3. Sugimoto CR, Larivière V. Measuring Research: What Everyone Needs to Know. Oxford: Oxford University Press; 2018. ISBN: 9780190640125. https://global.oup.com/academic/product/measuring-research-9780190640125

4. Ding Y, Rousseau R, Wolfram D. Measuring Scholarly Impact: Methods and Practice. Cham: Springer; 2014. DOI: 10.1007/978-3-319-10377-8

5. Börner K, Chen C, Boyack KW. Visualizing Knowledge Domains. In: Cronin B, Sugimoto CR, editors. Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact. Cambridge (MA): MIT Press; 2014. p. 197–228.DOI: 10.7551/mitpress/9780262026792.003.0010

6. Chen C. CiteSpace: A Practical Guide for Mapping Scientific Literature. Hauppauge (NY): Nova Science Publishers; 2016. ISBN: 9781634842847

BIBLIOGRAPHIC REFERENCES

1. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 1 de noviembre de 2017;11(4):959-75.

2. Pereira V, Basilio MP, Santos CHT. PyBibX – a Python library for bibliometric and scientometric analysis powered with artificial intelligence tools. Data Technologies and Applications. 2025;59(2):302-37.