Journal of Informetrics 20 (2026) 101837

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Research Paper

Biblioshiny and the SAAS Workflow: An integrated framework for transparent and reproducible science mapping. A demonstration through the replication of a study

Massimo Aria [a,b,*], Corrado Cuccurullo [a,b,c], Luca D'Aniello [a,b], Maria Spano [a,b]

[a] Department of Economics and Statistics, University of Naples Federico II, Naples, Italy
[b] K-Synth academic Spinoff, University of Naples Federico II, Naples, Italy
[c] Department of Economics and Management, University of Campania Luigi Vanvitelli, Capua, Italy

A R T I C L E I N F O

Keywords:
Biblioshiny
Bibliometrix
SAAS workflow
Science mapping
Reference disambiguation
Life cycle analysis
Thematic map

A B S T R A C T

The rapid expansion of bibliometric tools and methods has increased analytical opportunities but also generated fragmented, non-standardized workflows that undermine transparency and replicability. Our paper presents Biblioshiny, the graphical interface of the *Bibliometrix* R package, and its integrated SAAS (Search-Appraisal-Analysis-Synthesis) workflow for transparent and reproducible science mapping. The workflow unifies all stages of bibliometric research within a single, parameter-documented environment. To demonstrate its functionality, we apply it to the replication of Najmaei and Sadeghinejad's (2023) study on *Green and Sustainable Business Models*.

In the replication, automated reference matching identified 9748 variants (24.4% of total references). Our logistic growth updated forecast of field maturity, shifting the projected inflection point to 2027. Additionally, the co-word analysis identified consolidated thematic clusters as well as emerging research topics within the field.

Overall, the paper presents Biblioshiny and the SAAS workflow as an open, scalable, and auditable solution for bibliometric research, promoting cumulative validation, cross-study comparability, and wider accessibility for researchers, evaluators, and policy analysts.

1. Introduction

Bibliometrics and science mapping have become increasingly central in contemporary research evaluation and knowledge synthesis. The exponential growth of scholarly output has produced an unprecedented information overload, which in turn requires systematic tools to filter, organize, and interpret research trends (Eppler & Mengis, 2004; Shahrzadi et al., 2024). At the same time, rising demands for accountability in the allocation of resources and the broader agenda of open science have positioned bibliometric methods as indispensable instruments for both academic and policy communities (Matorevhu, 2024; van Raan, 1996).

While bibliometrics has traditionally been applied to performance appraisal, assessing researchers, institutions, and journals through citation-based indicators, (Bota-Avram, 2023; Cuccurullo et al., 2016), this evaluative use has been widely criticized for its

* Corresponding author.
E-mail addresses: massimo.aria@unina.it (M. Aria), corrado.cuccurullo@unicampania.it (C. Cuccurullo), luca.daniello@unina.it (L. D'Aniello), maria.spano@unina.it (M. Spano).

https://doi.org/10.1016/j.joi.2026.101837
Received 23 February 2026; Received in revised form 18 May 2026; Accepted 25 May 2026

reliance on reductionist metrics and its vulnerability to misuse (Bornmann & Marx, 2014; Gingras, 2016; Põder, 2022). Initiatives such as the San Francisco Declaration on Research Assessment (DORA) and the Coalition for Advancing Research Assessment (COARA) have explicitly advocated for more balanced and responsible approaches (Cagan, 2013; Gärtner et al., 2024).

Beyond evaluation, bibliometrics also serves the complementary purpose of science mapping, that is, the quantitative extraction and visualization of the intellectual, social, and conceptual structures of scientific domains (Aria & Cuccurullo, 2017, 2026; Börner et al., 2003; Moral-Muñoz et al., 2019, 2020; Small, 1999). Science mapping is particularly valuable because it provides a robust quantitative backbone to various forms of literature reviews by helping scholars identify key actors, thematic clusters, and research trajectories in a transparent and replicable manner.

Moreover, the field is currently characterized by a rapid proliferation of bibliometric tools. While this growth has widened the availability of methods and interfaces, ranging from proprietary platforms (e.g., Web of Science, Scopus) to open-source solutions (e.g., VOSviewer, CiteSpace, Bibliometrix), it has also produced a highly fragmented ecosystem. Tools vary considerably in terms of data coverage (e.g., inclusion of different databases and metadata fields), analytical functionality (e.g., network construction, clustering algorithms, statistical models), and usability (graphical interfaces versus code-based environments) (Bales et al., 2020; Bankar & Lihitkar, 2019; Tomaszewski, 2023).

Despite this diversity, current practice is still dominated by tool-driven and idiosyncratic procedures, where researchers combine partial functionalities across platforms without an explicit, end-to-end workflow that connects search decisions to analytical steps and, ultimately, to synthesis. As a result, "how" a bibliometric study has been conducted often remains implicit: crucial choices (e.g., query design, screening rules, inclusion thresholds, normalization options, clustering settings, and visualization parameters) are either under-reported or reported in a non-comparable way across studies. This weakens reproducibility and interpretability, not because bibliometrics lacks methods, but because the methodological logic that links data acquisition, appraisal, analysis, and synthesis is rarely operationalized in a consistent and auditable sequence.

To address this gap, this paper positions Biblioshiny as a workflow-compliant software environment for bibliometric and science mapping studies. Specifically, we introduce the SAAS workflow (Search-Appraisal-Analysis-Synthesis), a reformulation of Grant and Booth's (2009) SALSA framework (Search-Appraisal-Synthesis-Analysis) that reorders and redefines the stages to embed explicit parameter documentation and reproducibility requirements at each step, as a structured process model and show how it is implemented step-by-step within Biblioshiny, the graphical interface of the R-based Bibliometrix package (Aria & Cuccurullo, 2017, 2026). Rather than proposing yet another analytical technique or offering a purely conceptual framework, we focus on the software-level operationalization of SAAS: a guided pipeline that supports traceable decisions, standardized reporting of key parameters, and replicable outputs across the entire research cycle.

RQ1. How does Biblioshiny implement the SAAS workflow as a coherent analytical pipeline and enable the structured documentation and reporting of methodological choices to support transparency and comparability?

The primary scientific contribution of this paper is methodological: the SAAS framework as a principled approach to transparent, reproducible science mapping. Biblioshiny serves as the empirical vehicle through which this contribution is demonstrated and validated, offering a concrete instantiation of SAAS principles within an operational environment. Building on this premise, the empirical component of the paper shows how a SAAS-compliant implementation supports (i) the systematic capture and reporting of methodological decisions and (ii) the reproducibility of science mapping outputs under equivalent data sources and parameter settings. Within this perspective, replication exercises are interpreted as empirical validation of the workflow itself, rather than as a performance comparison between software tools. In the same spirit, comparative analyses are framed as a contrast between integrated workflows and fragmented practices, not as a competition between Biblioshiny and alternative platforms.

RQ2. How does a SAAS-compliant implementation in Biblioshiny affect the replicability of results and the accessibility of science mapping for non-programming users, under equivalent data sources and parameter choices?

By framing Biblioshiny as SAAS-compliant software, the paper contributes a practical, end-to-end solution to fragmentation: it makes the workflow explicit, the decisions auditable, and the outputs more reproducible and interpretable.

2. Background

Bibliometrics has long been established as a quantitative approach to the study of science, encompassing a wide range of analytical techniques such as productivity analysis, bibliographic coupling, citation and co-citation analysis, co-word analysis, and the mapping of collaboration networks (Klarin, 2024). These methods enable scholars to identify the structure and dynamics of scientific fields, providing insights into research productivity, intellectual influences, and emerging themes (Ciampi et al., 2021). Despite their widespread use, these approaches are subject to well-known limitations, including coverage bias across databases, non-normalized metrics that hinder comparability across fields, and a lack of transparency regarding parameter settings and methodological choices.

The increasing complexity of bibliometric research has been accompanied by the proliferation of dedicated tools and platforms, both proprietary and open-source. Major bibliographic databases such as *Web of Science, Scopus, Dimensions*, and *OpenAlex* provide the raw data, while a diverse ecosystem of software, ranging from *VOSviewer, SciMAT,* and *CiteSpace* to *Bibliometrix* and its graphical interface *Biblioshiny*, enables science mapping and performance analyses (Chen, 2016; Cobo et al., 2011; Cobo et al., 2012; Moral-Muñoz et al., 2020; Van Eck & Waltman, 2010). Yet this diversity also entails significant challenges. Bibliographic databases are often affected by limited accessibility and by a lack of full interoperability across platforms. At the same time, many bibliometric tools suffer from a lack of algorithmic transparency, frequently operating as *black boxes*, as well as from incomplete analytical coverage. In addition, persistent usability issues either confine their use to technically skilled users or, conversely, oversimplify inherently complex analytical processes (Gasparyan, 2013; Moral-Muñoz et al., 2020).

Bibliometrix (Aria & Cuccurullo, 2017, 2026) is today regarded as one of the most widely used tools for bibliometric analysis. It is an open-source R package that provides researchers with computational expertise an exceptionally high degree of flexibility, allowing them to design bibliometric studies with full control over both analytical procedures and graphical outputs. Many bibliometric tools differ in the extent to which analytical parameters are user-visible and adjustable; Bibliometrix allows users to explicitly define and modify relevant parameters. A practical consideration is that using the package typically requires familiarity with R, which may affect adoption across user groups. To address this gap and broaden accessibility, we, the authors of Bibliometrix, developed a companion application with a user-friendly GUI: Biblioshiny.

It, like its parent package Bibliometrix, can serve the dual purpose of performance appraisal and science mapping in support of literature reviews. However, it has been specifically designed with the latter objective in mind. To this end, Biblioshiny follows our proposed SAAS workflow (Search-Appraisal-Analysis-Synthesis), that is our reformulation of the seminal SALSA model (Search - Appraisal - Synthesis - Analysis), which systematizes the process of evidence collection and synthesis (Grant & Booth, 2009). The SAAS reformulation is a specific contribution of this paper: while SALSA structures the evidence synthesis process, SAAS extends it by embedding explicit parameter documentation and reproducible reporting requirements at each workflow stage, enabling auditable and comparable bibliometric studies.

Although previous studies have examined individual tools and methods (e.g., Kumar, 2025; Moral-Muñoz et al., 2020; Tomaszewski, 2023), there remains a lack of studies that evaluate bibliometric workflows with explicit attention to parameterization and replicability. This paper addresses that gap by presenting the SAAS workflow as implemented in Biblioshiny, illustrating its capacity to provide a coherent, transparent, and replicable approach to science mapping.

3. Biblioshiny and the SAAS workflow

Biblioshiny is a comprehensive web-based graphical user interface built on the R Shiny framework that serves as the primary interface for the Bibliometrix package. As a full-stack application, Biblioshiny operates through an interactive client-server architecture where the browser-based frontend communicates in real-time with R computational engines on the backend, enabling immediate visualization of analytical results without requiring users to write code or manage R sessions directly.

Biblioshiny's architecture leverages the reactive programming model inherent to Shiny, where user interactions with interface elements (e.g., parameter adjustments, filter selections, algorithm choices) trigger immediate recomputation of dependent analyses and automatic updating of visualizations. This real-time feedback mechanism enables exploratory data analysis workflows where users can iteratively refine parameters and observe their effects on results without manual re-execution steps. All computational operations are performed using the underlying Bibliometrix R functions, ensuring that analyses conducted through the graphical interface are

Table 1
Biblioshiny technical specifications.

Specification	Details
Software Type	Web-based graphical user interface (GUI) for bibliometric analysis
Parent Package	Bibliometrix (R package for comprehensive science mapping)
Architecture	Client-server reactive web application
Framework	R Shiny (v1.7+) with reactive programming model
Programming Language	R/CSS (backend), R/JavaScript/HTML/CSS (frontend)
Computational Engine	R statistical computing environment (v4.0+)
Platform Compatibility	Cross-platform (Windows, macOS, Linux); browser-based interface
License	Open source (MIT license)
Code Repository	https://github.com/massimoaria/bibliometrix
Installation	Via R: install.packages("bibliometrix") or CRAN (see Appendix A)
System Requirements	R $\geq$ 4.0.0; recommended RAM $\geq$ 8GB for large datasets (>10,000 documents)
Supported Browsers	Chrome, Firefox, Safari, Edge
Data Source Support	Web of Science, Scopus, Dimensions, PubMed, OpenAlex, Lens, Cochrane
Import Formats	.bib, .csv, .txt, RIS, JSON (via API integrations)
API Integration	Direct retrieval from OpenAlex, PubMed, Dimensions APIs
Maximum Dataset Size	Limited by available RAM; tested up to 100,000+ documents
Export Formats	Excel (.xlsx), CSV, PNG/SVG graphics, Network files (Gephi, Pajek), R data
Network Visualization	Interactive networks with force-directed layouts (D3.js), customizable node/edge styling
Analysis Modules	Data import/cleaning, Filters, Overview, Sources, Authors, Documents, Clustering, Conceptual/Intellectual/Social Structure, Report generation
AI/LLM Integration	Google Gemini API for automated interpretation support (optional)
Reproducibility Features	Parameter documentation, Session state export, R code generation, Configuration file sharing
Real-time Processing	Reactive updates on parameter changes; asynchronous processing for large datasets
Multilingual Support	Interface: English; Data: handles multiple character encodings (UTF-8, Latin-1)
Documentation	Integrated help system, tooltips, method references, video tutorials at https://bibliometrix.org
Update Frequency	Regular updates with new features and bug fixes; version-controlled releases
Community Support	GitHub issues, user forum, email support, active development community

methodologically equivalent to those performed via R scripts, thereby supporting reproducibility across different user interfaces. Biblioshiny achieves this without requiring programming skills: the entire SAAS pipeline is executable through a point-and-click interface, making workflow-compliant bibliometrics accessible to researchers regardless of their coding background (Aria & Cuccurullo, 2026).

From a usability perspective, Biblioshiny implements progressive disclosure of complexity: basic analyses require only dataset upload and minimal configuration, while advanced users can access extensive parameter controls, custom algorithm implementations,

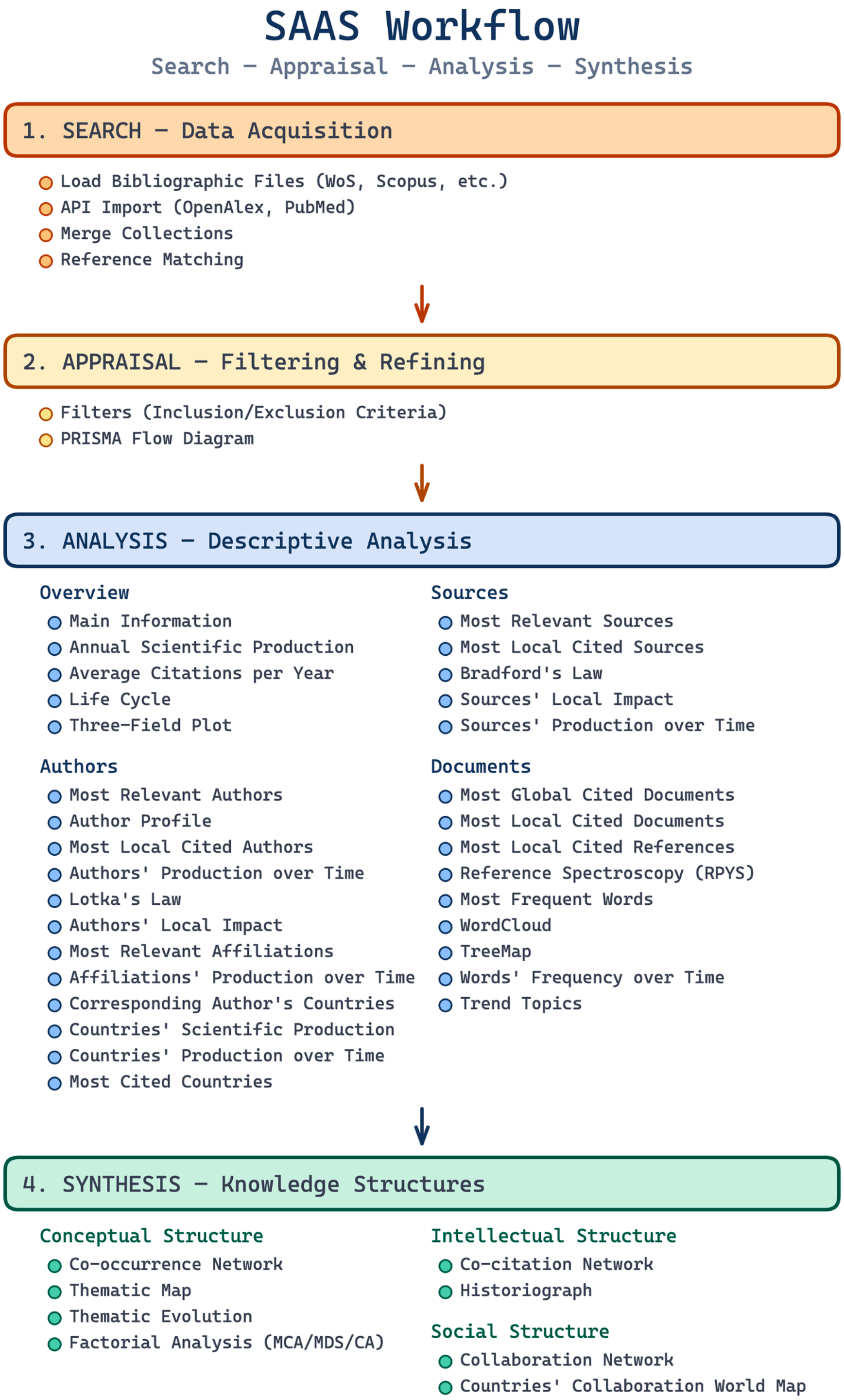


Fig. 1. The SAAS workflow as implemented in Biblioshiny.

and sophisticated visualization options through expandable interface sections. All parameter choices are preserved in session state and can be exported as configuration files for reproducibility or shared with collaborators to ensure consistent analytical settings across distributed research teams.

Table 1 provides a technical overview of Biblioshiny's key specifications and capabilities, while the process for install and run biblioshiny app is described in Appendix A.

Fig. 1 provides a comprehensive overview of how Biblioshiny operationalizes the SAAS workflow, mapping each analytical function in the four sequential stages of Search, Appraisal, Analysis, and Synthesis. The diagram highlights how data acquisition tools, filtering procedures, descriptive analyses, and knowledge structure techniques are integrated within a single coherent environment, thereby enabling the end-to-end execution of a science mapping study without recourse to external software.

3.1. Search

The Search phase of the SAAS workflow encompasses data acquisition, integration, and preparation for analysis. Biblioshiny offers robust capabilities for importing and harmonizing datasets from a wide range of bibliographic sources, including Web of Science, Scopus, Dimensions, PubMed, OpenAlex, and Lens. These sources can be accessed via standard bibliographic formats (e.g., *.bib, .csv*) or through direct API integration, depending on database licensing models (Velez-Estevez et al., 2023).

Biblioshiny provides a single interface where users can upload files, connect to APIs, merge datasets, and run cleaning routines within the 'Data' module. The merging functionality is particularly valuable for combining datasets from multiple sources, enhancing coverage while preserving metadata integrity and minimizing source-specific biases.

Among the available databases, Web of Science provides the most standardized metadata, improving data quality and consistency in downstream analyses (Birkle et al., 2020). Conversely, Google Scholar, often mistaken for a bibliographic database, is excluded due to its lack of structured metadata and reproducible export features (Boeker et al., 2013).

Overall, this integrated search environment allows for comprehensive and interoperable bibliometric investigations, ensuring transparency, reproducibility, and representativeness in large-scale science mapping studies.

Biblioshiny also embeds a comprehensive data cleaning and pre-processing layer, which operates both automatically and interactively. Duplicate records from single-source datasets are automatically detected and removed, while deduplication across merged datasets is managed algorithmically. The platform corrects inconsistent author and journal names and offers a semi-automated cited reference disambiguation system. Approximately 90% of reference variants are matched automatically; unresolved cases can be reviewed and reconciled through the interface. The same pre-processing outputs can then be reused across subsequent analytical steps within the workflow.

Keyword normalization is handled via the *Filters* section, where users can upload custom thesauri to consolidate plurals, synonyms, and spelling variants. This ensures conceptual coherence and enhances the quality of co-word and thematic analyses.

By consolidating all pre-processing operations into a single, reproducible pipeline, Biblioshiny consolidates pre-processing operations into a reproducible pipeline and makes intermediate steps and parameter settings easier to document.

3.2. Appraisal

The Appraisal phase establishes the inclusion and exclusion criteria that define the boundaries of the dataset and, by extension, the scope, relevance, and validity of the bibliometric analysis. Initial filtering typically relies on structural parameters derived from the search query, such as publication year, language, document type, and subject area. Biblioshiny further enables dataset refinement by applying more granular filters, including geographical focus, journal selection, and document categories (e.g., articles, reviews, proceedings).

Beyond structural attributes, appraisal can incorporate content- and method-based criteria, allowing users to prioritize studies that meet specific methodological standards, offer particular types of findings, or align closely with the research question. To support this process, Biblioshiny provides a dedicated *Appraisal* module that enables systematic filtering across authors, journals, and documents. Researchers can also upload custom lists of keywords or terms to enforce domain-specific screening criteria.

Importantly, Biblioshiny's appraisal functionality promotes transparency and reproducibility, ensuring that dataset selection is not left to ad hoc or manual procedures. The platform also includes strategy recommendations aligned with distinct research objectives, such as domain-level knowledge mapping, longitudinal journal analysis, or institutional performance assessment, thus enabling tailored yet replicable study designs.

3.3. Analysis

The Analysis phase is structured across three interconnected levels, each contributing to a comprehensive understanding of the scientific landscape under investigation.

3.3.1. Overview

This level provides a general snapshot of the dataset through key indicators associated with journals, authors, and documents. Biblioshiny's dashboard offers visual displays of publication trends over time (e.g., annual growth of scientific output), performance summaries by field (*three-field plots*), and logistic curve analysis (LCA) to model the life cycle of the research domain. These indicators serve as a preliminary orientation, supporting the identification of dominant actors and temporal dynamics.

3.3.2. Focus on domain (J-AU-DOC metrics)

This level delivers a more granular examination of scholarly activity across three core units of analysis: journals (J), authors (AU), and documents (DOC). At the journal level, productivity, impact measures, and Bradford's law highlight core outlets and dispersion patterns. At the author level, indices such as the h-index and g-index, Lotka's law, and collaboration metrics (e.g., co-authorship networks) provide insight into scholarly influence and team dynamics. At the document level, citation analyses - both raw and normalized - are complemented by robust metrics that account for skewed distributions, identifying high-impact contributions and emerging works.

3.3.3. Knowledge structures

Moving beyond descriptive analytics, this level explores the intellectual, social, and conceptual structures of the field. The first structure is reconstructed through co-citation analysis and direct citation analysis such as the Histograph. The social structure is mapped through co-authorship networks and institutional or international collaboration patterns, with centrality and modularity indices providing network-level insights. The conceptual structure is captured via co-word analysis, factorial-based analysis and clustering, or thematic diagrams, producing strategic and evolutionary maps that visualize the thematic organization and temporal evolution of research areas.

Together, these levels of analysis offer a multidimensional perspective that bridges performance evaluation, network dynamics, and knowledge mapping within a unified bibliometric framework.

3.4. Synthesis

Finally, the Synthesis phase integrates the results of bibliometric analysis into coherent representations that facilitate interpretation and reporting. Alongside classical statistical outputs, such as summary tables, frequency distributions, and descriptive indicators, bibliometrics is distinguished by its use of graphical syntheses that capture the relational and dynamic dimensions of science. These include network visualizations of co-authorship, co-citation, or co-word relationships; strategic diagrams and matrices that position thematic clusters according to centrality and density; and maps of knowledge domains that reveal intellectual, social, and conceptual structures. Longitudinal perspectives are provided through diachronic series and topic evolution maps, which trace the emergence, consolidation, or decline of research themes over time. Importantly, this phase also requires standardized reporting practices to ensure that parameter settings, methodological choices, and visualization strategies are transparently documented, thereby enhancing comparability and reproducibility across studies (Gan et al., 2022).

In addition to its structured organization according to the SAAS workflow, Biblioshiny integrates a wide range of basic and advanced functionalities that enhance its analytical capacity, user experience, and scientific rigor.

3.5. Basic functionalities

At the most accessible level, every analysis in Biblioshiny is provided in both graphical and data formats, with the option to download outputs for further elaboration. Data tables can be exported directly to Excel, ensuring interoperability with standard office and statistical software.

Users can configure a broad range of parameters for analyses and visualizations, which provide flexibility while ensuring methodological transparency. Thresholds, such as minimum word occurrences, author productivity levels, or citation counts, may be specified to prune sparse networks and enhance interpretability. The choice of normalization index (e.g., cosine similarity, Jaccard coefficient, Salton index, or equivalence index) can be tailored to the objectives of the study and the structure of the underlying data. For network construction, multiple clustering algorithms are available, including Louvain, Walktrap, Ward, and Leiden, each with tunable settings to optimize community detection (Singh & Garg, 2025). Similarly, layout configurations such as Fruchterman-Reingold, Kamada-Kawai, and related algorithms can be adjusted to balance visual stability and readability (Börner et al., 2007). Together, these options ensure that results are not only analytically rigorous but also reproducible and visually interpretable. A comprehensive overview of Biblioshiny's SAAS implementation and detailed parameter settings is provided in Appendix B; Tables B.1 and B.2.

Each output is accompanied by explanations and references to facilitate correct interpretation. Moreover, all results, both graphical and numerical, can be stored in a downloadable report, allowing for cumulative documentation of the analysis process. Hyperlinks to the DOIs of articles are provided whenever available, enabling direct access to the underlying publications.

3.6. Advanced functionalities

Beyond these baseline features, Biblioshiny incorporates advanced options that expand the scope of bibliometric research. The software allows for API integration, both for *Search* (e.g., PubMed, OpenAlex) and for AI/ML applications. In particular, large language models such as Google Gemini can be used to generate interpretive support for graphical and statistical outputs. Users can also upload thesauri, stoplists, and disambiguation files to improve consistency in both *Appraisal* and *Analysis* phases. Advanced modules include content analysis and in-context citation analysis, which enable researchers to classify citations rhetorically (e.g., background, gap, method, critique) and to integrate quantitative mapping with qualitative dimensions. Additional quality-control features address duplication checks and the disambiguation of authors, affiliations, and keywords. Recent tools such as BiblioMerge (Diez-Junguitu & Pena-Cerezo, 2026) further extend this capability by providing automated author disambiguation across WoS and Scopus exports, compatible with the Biblioshiny workflow. Significantly extending this line of functionality, Biblioshiny now natively supports the automated merging of collections retrieved from all seven bibliographic databases it handles: records from heterogeneous sources are normalized, de-duplicated, and converted into the standardized bibliometrix data format, enabling seamless integration of multi-source corpora within a single, coherent analytical pipeline.

Analytical power is further enhanced through the integration of statistical estimators and tests for power-law distributions (Lotka's law, Zipf's law, Bradford's law), normalized citation indicators, and author-level indices such as *h* and *g*.

4. Methods

This study constitutes a replication of the bibliometric article by Arash Najmaei and Zahra Sadeghinejad, "Green and Sustainable Business Models: Historical Roots, Growth Trajectory, Conceptual Architecture and an Agenda for Future Research" (Najmaei & Sadeghinejad, 2023). Our aim is to reproduce and extend its analyses by applying the SAAS workflow within the Biblioshiny interface. This study was selected as the replication target for three reasons: (i) the supplementary dataset (851 records) is publicly available, enabling independent verification; (ii) the original study employs a multi-tool, multi-step methodology (BibExcel, CRExplorer, LogletLab) that exemplifies the fragmented workflows SAAS aims to address; and (iii) its focus on Green and Sustainable Business Models provides a substantively relevant and well-bounded domain for science mapping.

4.1. Data collection and pre-processing (S + A)

In line with the original study, we did not perform independent data collection or define our own inclusion and exclusion criteria. Instead, we utilized the dataset of 851 records provided in the supplementary materials of the original article.

In the original work, data cleaning and pre-processing were performed manually over a period of approximately three weeks. This process involved two main activities: (1) identifying variants of cited references and (2) disambiguating authors' names and journal titles. These steps were conducted using multiple tools, such as BibExcel and CRExplorer, and required extensive manual verification. Notably, the disambiguated data were applied selectively - e.g., reference variants identified via CRExplorer were used exclusively for the RPYS (Referenced Publication Year Spectroscopy) analysis and were not exportable or reusable across other analyses relying on cited references.

In our replication, data cleaning was performed using Biblioshiny's automated routines, which operate across units of analysis (journals, authors, documents) and produce normalized metadata for subsequent steps. These transformations not only enhance data consistency but also facilitate the merging of datasets from multiple sources (e.g., Web of Science, Scopus, OpenAlex), thereby increasing interoperability and representativeness.

With regard to cited references, the original study manually identified approximately 200 variants (to be confirmed), whereas the automated reference matching algorithm in Biblioshiny identified 9.748 variants. A single reference, Osterwalder and Pigneur (2010), was associated with 116 variants, illustrating the extent of citation-format heterogeneity in the dataset.

5. Reference matching algorithm

Citation matching represents a persistent methodological challenge in bibliometric research, as inconsistencies in citation formats, punctuation, abbreviations, and incomplete metadata often inflate citation counts and compromise the accuracy of scholarly impact metrics (Leydesdorff et al., 2016; Pasula et al., 2002). To address this, the proposed methodology implements a six-step normalization and matching algorithm that combines deterministic and probabilistic approaches -format detection, ISO standardization, exact and fuzzy matching, blocking, and metadata-based consolidation - to maximize both precision and computational efficiency.

- **Step 1 - Format Detection and Metadata Extraction.** The algorithm begins by identifying the citation source format (Web of Science or Scopus) and extracting structured metadata using regular expressions. Text normalization follows established record linkage standards (Christen, 2012), including diacritic removal, case standardization, punctuation cleaning, and validation of key fields such as author, year, journal, volume, pages, DOI, and title keywords.
- **Step 2 - Journal Name Normalization.** Journal names are standardized to their ISO 4 abbreviated forms using the LTWA database maintained by ISO, thereby minimizing ambiguity caused by inconsistent or multilingual abbreviations (Beagrie & Houghton, 2014; Mongeon & Paul-Hus, 2016). This step ensures uniformity in subsequent string comparisons.
- **Step 3 - Exact Matching.** Deterministic matching first clusters citations with identical DOIs, treated as unique and persistent identifiers (Paskin, 2005), followed by string-based equality checks on normalized citation strings. This stage captures trivial variants prior to more computationally demanding fuzzy matching (Fellegi & Sunter, 1969).
- **Step 4 - Blocking Strategy.** To enhance scalability, the algorithm applies a blocking mechanism that limits comparisons to subsets sharing the same author, year, and partial journal key, reducing computational complexity from quadratic to quasi-linear (Baxter et al., 2003).
- **Step 5 - Within-Block Fuzzy Matching.** Within each block, matching is adapted to citation structure. Deterministic metadata-based grouping is applied to Web of Science data, while Scopus or mixed-format citations are matched through fuzzy string similarity using the Jaro-Winkler metric (Cohen et al., 2003; Winkler, 1990). Hierarchical clustering merges near-identical citations above a defined similarity threshold (Aksnes et al., 2019; Bilenko et al., 2003).

- **Step 6 - Post-Processing and Canonicalization.** A final global merge corrects block-boundary effects and unifies citations with equivalent core metadata (Christen & Goiser, 2007; Leydesdorff et al., 2016). For each cluster, the algorithm selects the most complete citation as the canonical representative based on a weighted completeness score (Zhu et al., 2015) defined as: S = w1*DOI + w2*(field_count/max_fields) + w3*(string_length/max_length) where DOI presence is weighted most heavily (w1 = 0.5), followed by bibliographic field completeness (w2 = 0.3) and normalised string length (w3 = 0.2). DOI presence is treated as the primary criterion given its role as a persistent unique identifier (Paskin, 2005).

5.1. Data analysis (A + S)

For the Life Cycle Analysis (LCA), while the original study employed *LogletLab 4.0*, our implementation used the logistic curve fitting module available in Biblioshiny, which allowed model estimation to be conducted within the same workflow environment.

Reference Publication Year Spectroscopy (RPYS) was applied to examine the temporal distribution of cited references within the dataset. The original study conducted this analysis using CRExplorer. In the present replication, RPYS was implemented through the "Focus on Domain" module available in Biblioshiny, which generates a spectrogram of citation frequencies over time and identifies peaks associated with highly cited reference years. While the original article did not report detailed analytical or graphical parameter settings, the procedure adopted here was executed with parameter configurations that can be inspected and documented within the workflow environment.

Additionally, Biblioshiny automatically categorizes references into four distinct classes based on their temporal citation patterns, following the classification framework proposed by Thor et al. (2018). This categorization is performed through automated analysis of citation trajectories over time, enabling the identification of different types of influential publications within the field. The four categories are: (1) *Hot Papers*: publications experiencing a recent surge in citations, indicating current relevance and impact in the field. These works are often at the forefront of emerging trends or active debates; (2) *Life Cycle Papers*: publications that have maintained a consistent level of citations over an extended period, demonstrating enduring foundational importance and continued relevance to the research field. These are typically considered classics within the domain; (3) *Sleeping Beauties*: publications that received limited initial attention but have experienced a significant increase in citations over time, indicating delayed recognition of their impact or renewed interest in their contributions (Ke et al., 2015; van Raan, 2004); (4) *Constant Performers*: publications with consistently high citation rates over time, suggesting they are continuously referenced by researchers. These works represent stable and established knowledge within the field.

This automated classification provides insight into the temporal dynamics of scholarly influence, complementing the RPYS analysis by revealing not only which years are heavily cited but also the citation patterns that characterize foundational versus emerging contributions.

To explore the conceptual structure of the field, a co-word analysis was conducted using the *Conceptual Structure* function in Biblioshiny. The analysis focused on author keywords (DE field) and covered two distinct periods (2000–2013 and 2014–2022) to capture the evolution of thematic patterns over time. The 250 most frequent keywords (*n* = *250*) were used, with a minimum frequency threshold of five occurrences (*minFreq* = *5*) to ensure statistical robustness while filtering out infrequent terms. A single-word

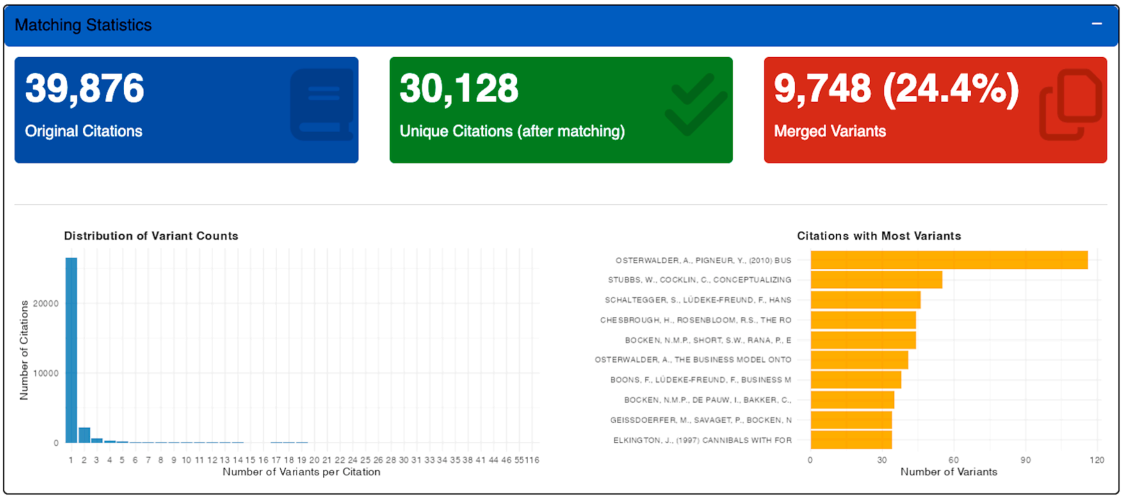
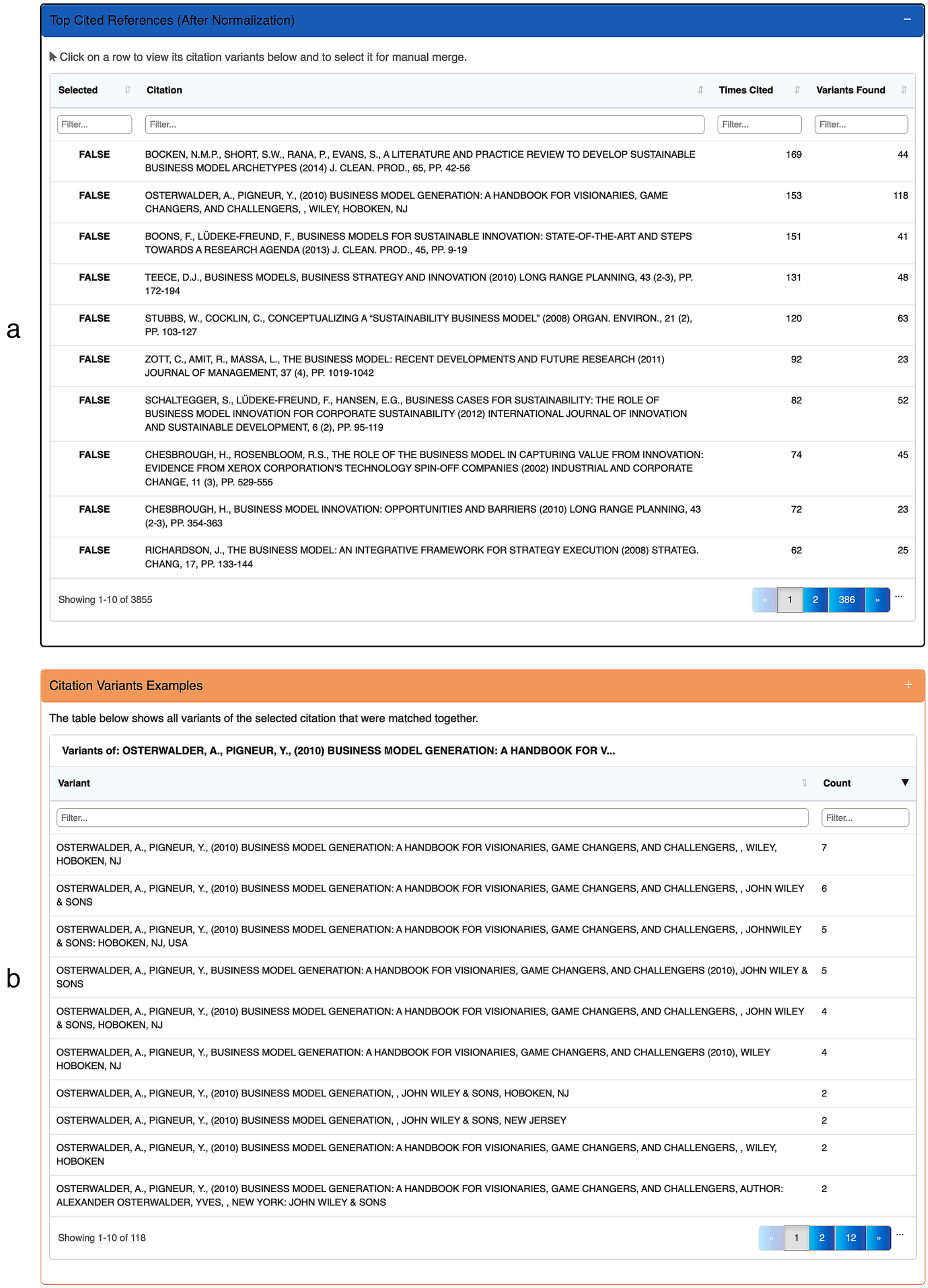


Fig. 2. Automated cited reference matching results.
Distribution of citation variants identified through the automated reference matching algorithm. The top panel shows the total number of raw citations (Original Citations) versus unique citations (after matching), with 9748 merged variants representing 24.4% of the total dataset. The bottom-left panel displays the frequency distribution of variant counts per reference, revealing a heavily right-skewed pattern. The bottom-right panel lists the top-10 most fragmented citation



Selected	Citation	Times Cited	Variants Found
FALSE	BOCKEN, N.M.P., SHORT, S.W., RANA, P., EVANS, S., A LITERATURE AND PRACTICE REVIEW TO DEVELOP SUSTAINABLE BUSINESS MODEL ARCHETYPES (2014) J. CLEAN. PROD., 65, PP. 42-56	169	44
FALSE	OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , WILEY, HOBOKEN, NJ	153	118
FALSE	BOONS, F., LÜDEKE-FREUND, F., BUSINESS MODELS FOR SUSTAINABLE INNOVATION: STATE-OF-THE-ART AND STEPS TOWARDS A RESEARCH AGENDA (2013) J. CLEAN. PROD., 45, PP. 9-19	151	41
FALSE	TEECE, D.J., BUSINESS MODELS, BUSINESS STRATEGY AND INNOVATION (2010) LONG RANGE PLANNING, 43 (2-3), PP. 172-194	131	48
FALSE	STUBBS, W., COCKLIN, C., CONCEPTUALIZING A “SUSTAINABILITY BUSINESS MODEL” (2008) ORGAN. ENVIRON., 21 (2), PP. 103-127	120	63
FALSE	ZOTT, C., AMIT, R., MASSA, L., THE BUSINESS MODEL: RECENT DEVELOPMENTS AND FUTURE RESEARCH (2011) JOURNAL OF MANAGEMENT, 37 (4), PP. 1019-1042	92	23
FALSE	SCHALTEGGER, S., LÜDEKE-FREUND, F., HANSEN, E.G., BUSINESS CASES FOR SUSTAINABILITY: THE ROLE OF BUSINESS MODEL INNOVATION FOR CORPORATE SUSTAINABILITY (2012) INTERNATIONAL JOURNAL OF INNOVATION AND SUSTAINABLE DEVELOPMENT, 6 (2), PP. 95-119	82	52
FALSE	CHESBROUGH, H., ROSENBLOOM, R.S., THE ROLE OF THE BUSINESS MODEL IN CAPTURING VALUE FROM INNOVATION: EVIDENCE FROM XEROX CORPORATION'S TECHNOLOGY SPIN-OFF COMPANIES (2002) INDUSTRIAL AND CORPORATE CHANGE, 11 (3), PP. 529-555	74	45
FALSE	CHESBROUGH, H., BUSINESS MODEL INNOVATION: OPPORTUNITIES AND BARRIERS (2010) LONG RANGE PLANNING, 43 (2-3), PP. 354-363	72	23
FALSE	RICHARDSON, J., THE BUSINESS MODEL: AN INTEGRATIVE FRAMEWORK FOR STRATEGY EXECUTION (2008) STRATEG. CHANG, 17, PP. 133-144	62	25

Variant	Count
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , WILEY, HOBOKEN, NJ	7
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , JOHN WILEY & SONS	6
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , JOHNWILEY & SONS: HOBOKEN, NJ, USA	5
OSTERWALDER, A., PIGNEUR, Y., BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS (2010), JOHN WILEY & SONS	5
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , JOHN WILEY & SONS, HOBOKEN, NJ	4
OSTERWALDER, A., PIGNEUR, Y., BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS (2010), WILEY HOBOKEN, NJ	4
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION, , JOHN WILEY & SONS, HOBOKEN, NJ	2
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION, , JOHN WILEY & SONS, NEW JERSEY	2
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, , WILEY, HOBOKEN	2
OSTERWALDER, A., PIGNEUR, Y., (2010) BUSINESS MODEL GENERATION: A HANDBOOK FOR VISIONARIES, GAME CHANGERS, AND CHALLENGERS, AUTHOR: ALEXANDER OSTERWALDER, YVES, , NEW YORK: JOHN WILEY & SONS	2

Fig. 3. Citation variants analysis.
Detailed breakdown of citation variants for the most influential references in the dataset. Panel (a) presents part of the list of references ranked by total citations and number of associated variants, highlighting works with the highest degree of citation fragmentation. Panel (b) displays the specific variant strings identified for each canonical reference, illustrating the diversity of citation formats (abbreviations, punctuation, incomplete metadata) that contribute to bibliometric redundancy.

tokenization approach was adopted (*ngrams* = *1*), and stemming was disabled (*stemming* = *FALSE*) to preserve the semantic precision of domain-specific terminology. The node size parameter (*size* = *0.3*) determined the proportional scaling of keyword nodes, enhancing readability of prominent clusters. To maintain visual clarity, label density was limited to three per cluster (*n.labels* = *3*), and node repulsion was deactivated, allowing for a compact and interpretable layout.

For community detection, the Louvain clustering algorithm (Blondel et al., 2008) was applied to maximize modularity and identify cohesive thematic groupings within the keyword co-occurrence network. Following Newman and Girvan (2004), modularity values above the 0.3 threshold are generally considered indicative of a meaningful community structure.

6. Results

6.1. Precleaning: automated cited reference matching

To improve data quality and analytical reliability, the cited reference matching procedure described in the previous section was applied to the original dataset. Out of 39.876 raw citations, only 30.128 were identified as unique citations after the matching process, resulting in 9.748 merged variants, which represents 24.4% of the total (Fig. 2). This level of redundancy highlights a common issue in bibliometric datasets, where the same reference may appear in multiple syntactic variants due to inconsistencies in formatting, incomplete metadata, or citation errors.

In Fig. 2, the distribution of variant counts (bottom-left panel) reveals a heavily right-skewed pattern, where most citations have a low number of variants (typically one or two), but a small number of references are associated with a disproportionately high number of distinct variants. This phenomenon is especially evident in the top-10 most fragmented citations (bottom-right panel), which include foundational works such as Osterwalder & Pigneur's *Business Model Generation* and seminal contributions by authors like Stubbs, Cook, and Bocken. Notably, Osterwalder and Pigneur (2010) alone accounted for >110 variants, underscoring the importance of disambiguation procedures for accurate co-citation and network analyses.

The application of reference matching is therefore a crucial quality control step in bibliometric workflows, directly affecting the validity of intellectual structure mappings and the precision of citation-based indicators.

The analysis of citation variants reveals that some of the most influential references in the dataset are also those most affected by fragmentation. As shown in the Fig. 3, Osterwalder and Pigneur's (2010) *Business Model Generation* exhibits the highest degree of citation variability, with 116 distinct variants, despite being cited 151 times. Similarly, Stubbs and Cocklin (2008) and Schaltegger et al. (2012) each appear with over 40 variants, highlighting the risk of underestimating citation impact if variant matching is not performed. Other frequently cited works, such as Teece (2010); Zott et al. (2011), and Chesbrough and Rosenbloom (2002) contributions to business model innovation, also display substantial variant counts.

This phenomenon is particularly salient in fields like sustainable business models and business model innovation, where core references are widely disseminated across disciplines and publication outlets. The presence of dozens of variants for single works undermines the validity of citation metrics and co-citation analyses unless properly reconciled. These findings further reinforce the necessity of performing reference disambiguation as a prerequisite for robust bibliometric studies, particularly in science mapping applications where accurate network structures depend on consistent citation identification.

6.2. Sample overview and life cycle and growth dynamics

The dataset comprises 851 documents published between 2002 and 2019 across 481 distinct sources, reflecting an average annual growth rate of 27.83% - a clear indicator of a rapidly expanding field. In total, 808 unique authors contributed to this corpus, with an average of one author per document and an international co-authorship rate of 28.91%, suggesting a predominance of individual contributions alongside modest global collaboration. The dataset's 31.425 cited references and 2.158 author keywords reveal a rich and diverse thematic landscape (Fig. 4).

A logistic growth model was applied to assess the developmental stage of the research domain (Fig. 5). The model exhibits an excellent fit (R2 = 0.989; RMSE = 5.60, representing <1% of the projected annual peak production of 607 publications, confirming a high degree of model accuracy) and estimates a saturation point (K) of 8510 publications. The peak of annual production is projected for 2027, with approximately 607 publications per year, while the cumulative growth curve suggests a total expansion duration of about 15.4 years. It should be noted that these projections are estimates based on the trend observed up to 2019 (the dataset cutoff),

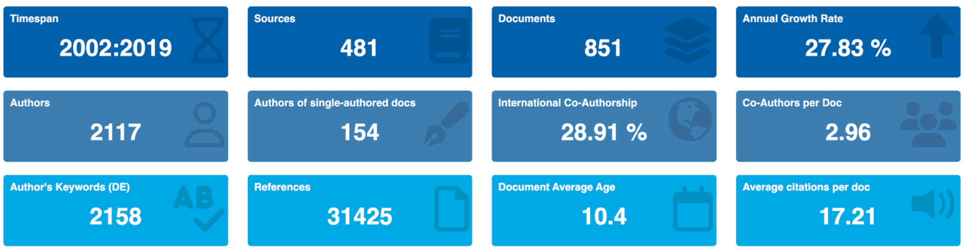


Fig. 4. Main information about the collection.

Fig. 5. Life cycle analysis of the research domain.

and may be sensitive to post-2019 developments in publication rates, funding patterns, or methodological shifts in the GSBM field. As of 2019, the field had achieved only 10% of its projected saturation, placing it firmly in the rapid growth phase. Forecasts indicate a continued upward trajectory, with around 490 new publications expected by 2024 and over 5.000 cumulative outputs by 2029. The resulting S-shaped diffusion pattern confirms the field's dynamism and underscores the value of ongoing science mapping analyses.

The life cycle trajectory reveals three main phases of development. The Emergence Phase (2000-ca. 2010) corresponds to low publication activity and the formative conceptualization of the domain. The Expansion Phase (ca. 2010–2027) is characterized by exponential growth, likely propelled by intensified academic engagement, expanding funding opportunities, and advances in methodological or technological tools. From 2027 onward, the field appears to enter a Maturity or Saturation Phase, as indicated by a gradual decline in annual publication growth and the flattening of the cumulative curve, signals of diminishing marginal returns in scholarly novelty.

From a scientific perspective, this logistic pattern suggests a transition from rapid expansion toward conceptual consolidation. The maturing trajectory implies that foundational questions have largely been established, with research now oriented toward incremental refinement rather than paradigm shifts. Declining growth rates may also reflect shifting funding priorities or resource reallocation toward adjacent emerging topics. Conversely, the earlier acceleration of output likely mirrored the influence of methodological and computational innovations that initially galvanized the field. The current plateau, therefore, may precede a new conceptual or methodological turning point, marking the onset of a fresh cycle of renewal.

6.3. Focus on domain and reference spectroscopy

The RPYS (Reference Publication Year Spectroscopy) plot (Fig. 6) offers a longitudinal perspective on the intellectual roots and developmental trajectory of the Green and Sustainable Business Models (GSBM) field. Foundational peaks in the early 1960s - particularly in 1962 and 1966 - indicate a grounding in environmental awareness, systems theory, and organizational strategy. A significant increase in cited references occurs between the late 1970s and early 1980s, corresponding to the emergence of corporate social responsibility (Carroll, 1979), competitive strategy (Porter, 1979), and stakeholder theory (Freeman, 2010). These works laid the groundwork for a more structured conceptualization of GSBM. The prominent spike in 1987 reflects the influence of the Brundtland Report (Our Common Future), which institutionalized the concept of sustainable development and shaped subsequent discourse.

The 1990s mark a period of theoretical consolidation, with frequent citations to resource-based view literature (Barney, 1991; Hart, 1995), stakeholder integration (Donaldson & Preston, 1995), and ecological modernization theory (Porter & van der Linde, 1995). During the early 2000s, the concept of the business model gains prominence, as evidenced by peaks associated with Chesbrough and Rosenbloom (2002); Magretta (2002), and Mont (2002), reflecting the strategic articulation of sustainability within innovation frameworks. From 2010 onwards, a sharp growth in citations signals rapid consolidation, driven by works on business model innovation (Chesbrough, 2010; Osterwalder & Pigneur, 2010; Teece, 2010), sustainable innovation (Boons et al., 2013), and circular economy frameworks (Bocken et al., 2014; Geissdoerfer et al., 2017).

Categorizing the most influential references further elucidates the structure of the field. "Constant performers" include Boons and Lüdeke-Freund (2013); Chesbrough (2010), and Bocken et al. (2014), which maintain high citation levels over time. "Hot papers," such as Wagner and Svensson (2010), exhibit recent spikes in attention, signaling their role in shaping current debates. "Sleeping beauties" (e.g., Belk, 2014; Richardson, 2008; Stubbs & Cocklin, 2008) show delayed yet increasing recognition, while certain publications like the Brundtland Report (1987); Tukker (2015), and Chesbrough and Rosenbloom (2002) display a "life cycle" pattern, with sustained influence across decades.

Nonetheless, the bibliometric analysis invites further critical reflection. Potential disciplinary biases may skew the representation of perspectives, particularly toward management and economics, while geographical asymmetries could indicate a dominance of Anglo-American scholarship. Methodological pluralism also warrants scrutiny: are diverse research designs and epistemologies

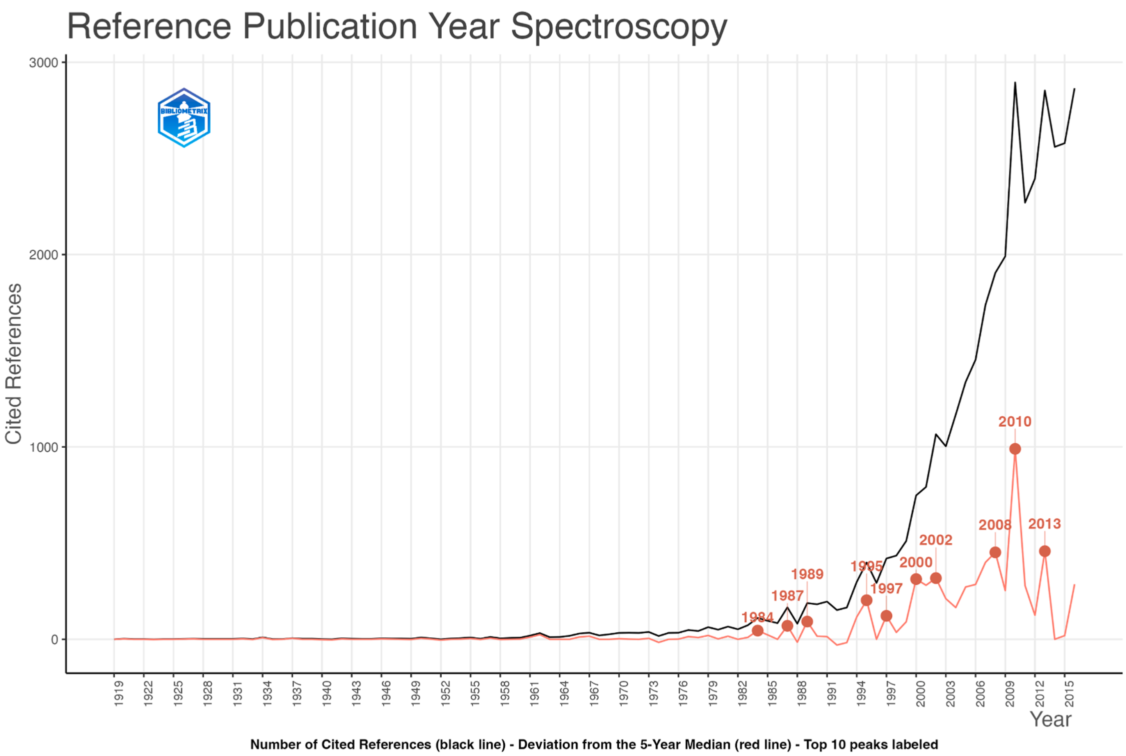


Fig. 6. Reference publication year spectroscopy (RPYS).

sufficiently represented? Furthermore, the nature of citation practices, whether substantive or perfunctory, should be interrogated to assess the true intellectual impact of seminal works. Finally, emerging trends, including circularity, social entrepreneurship, and AI-driven sustainability strategies, point to the evolving frontiers of GSBM research.

Altogether, this bibliometric profile not only reveals the interdisciplinary and temporal dynamics of the GSBM field but also underscores the need for reflective and critical engagement with its intellectual foundations and future directions.

6.4. Conceptual knowledge structure: thematic evolution and strategic positioning

The thematic evolution analysis, based on the Sankey diagram and strategic maps, highlights the progression and consolidation of research themes in the field of Green and Sustainable Business Models (GSBM) across two periods: 2002–2013 and 2014–2019.

The algorithm yielded a modularity score of Q = 0.7641 for the first period (2000–2013), well above this threshold and denoting a strongly partitioned network with well-separated thematic communities, and Q = 0.3319 for the second period (2014–2022), which only marginally exceeds the threshold and reflects a more interconnected and less sharply delineated conceptual structure. The resulting strategic diagrams and network maps plotted clusters according to centrality (inter-cluster relevance) and density (intra-cluster cohesion), following the classical Callon et al. (1991) framework. This enabled the classification of themes into motor, basic, niche, and emerging/declining categories, thereby illustrating the conceptual organization and longitudinal development of the field across the two periods.

The Sankey diagram (Fig. 7) reveals the continuity and transformation of core concepts. *Sustainability* emerges as a consistently central theme, maintaining strong linkages across both periods and absorbing several related strands such as *sustainable business models* and *base of the pyramid*. In contrast, themes like *e-business* and *sustainable development* appear to have been subsumed under broader constructs. Notably, *Circular Economy* gains prominence in the later period, indicating a strategic shift towards circularity in sustainable business discourse. The appearance of Telemedicine as a niche theme warrants critical reflection: it may reflect a genuine interdisciplinary convergence, digital health platforms increasingly adopting sustainable or circular business models, or it could be an artefact of keyword co-occurrence patterns amplified by the metadata normalization process. A qualitative review of the underlying documents would be necessary to discriminate between these two interpretations.

The strategic diagrams provide a more granular understanding of each theme's positioning in terms of centrality (relevance) and density (development). In the 2002–2013 period (Fig. 8), *sustainability, environmental sustainability, business models*, and *Internet innovation* are classified as motor themes, both central and well-developed. At the same time, *sustainable development, corporate sustainability*, and *e-health* emerge as basic themes, important but less mature. Topics such as *energy poverty, lean manufacturing*, and *health information exchange* appear as niche themes, while *sustainable innovation* and *economic sustainability* occupy the emerging or declining quadrant.

By contrast, the 2014–2019 map (Fig. 9) shows a more consolidated thematic structure. *Circular economy, business model innovation*, and *sustainable business models* cluster tightly as motor themes, reinforcing their central role in driving the field. *Industry 4.0* appears as

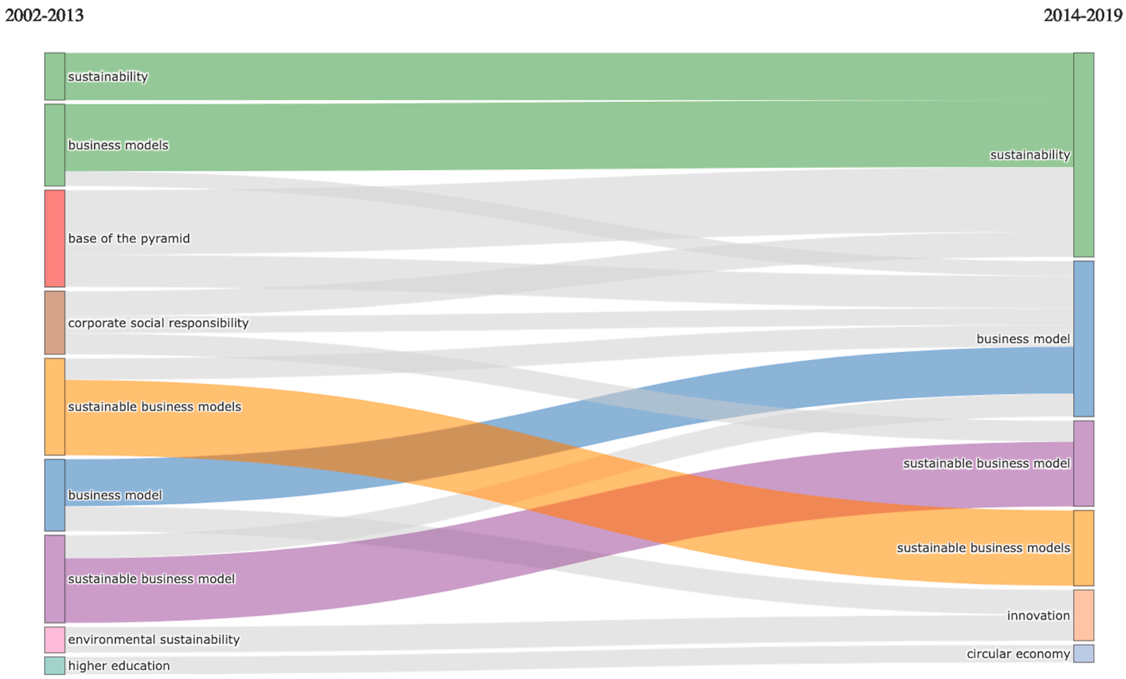


Fig. 7. Thematic evolution through Sankey diagram.

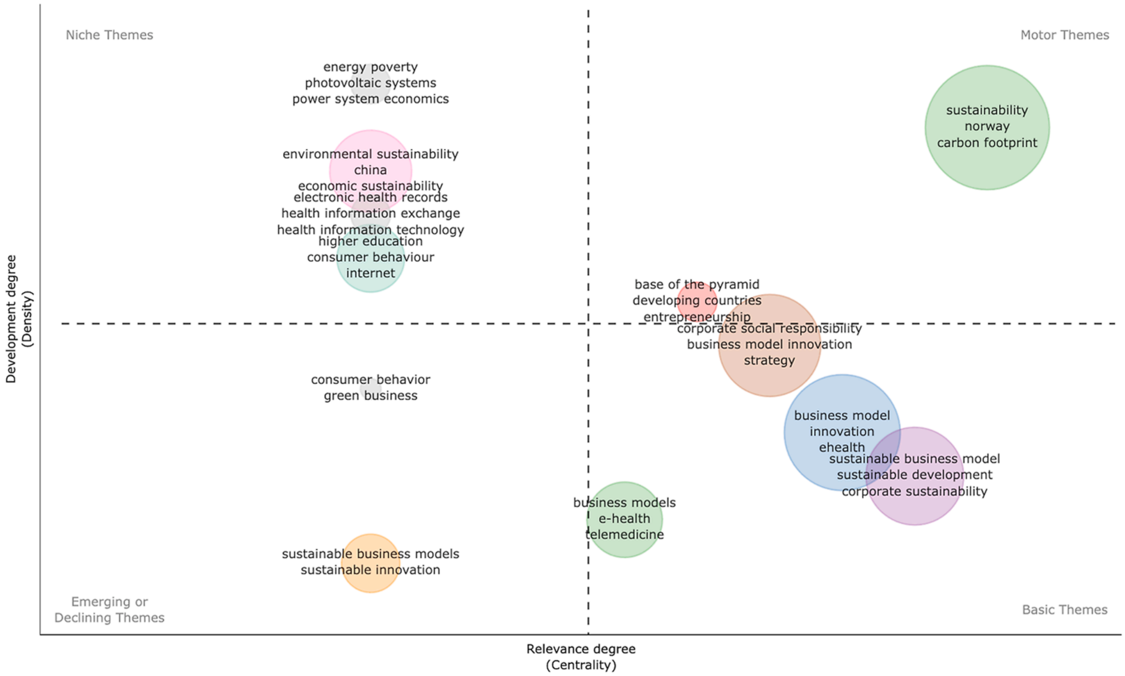


Fig. 8. Strategic thematic map – first period (2002–2013).

a basic theme, suggesting growing but still underdeveloped interest, while *environmental sustainability* and *telemedicine* are positioned as niche themes. Notably, no themes occupy the emerging/declining quadrant in this period, indicating a temporary stabilization of thematic exploration.

In synthesis, the field shows increasing consolidation around core themes like *sustainability* and *circular economy*, while simultaneously fostering specialized niches. The emergence of *Industry 4.0* and *telemedicine* indicates diversification in application domains. These findings underscore both the maturity and dynamism of the GSBM field.

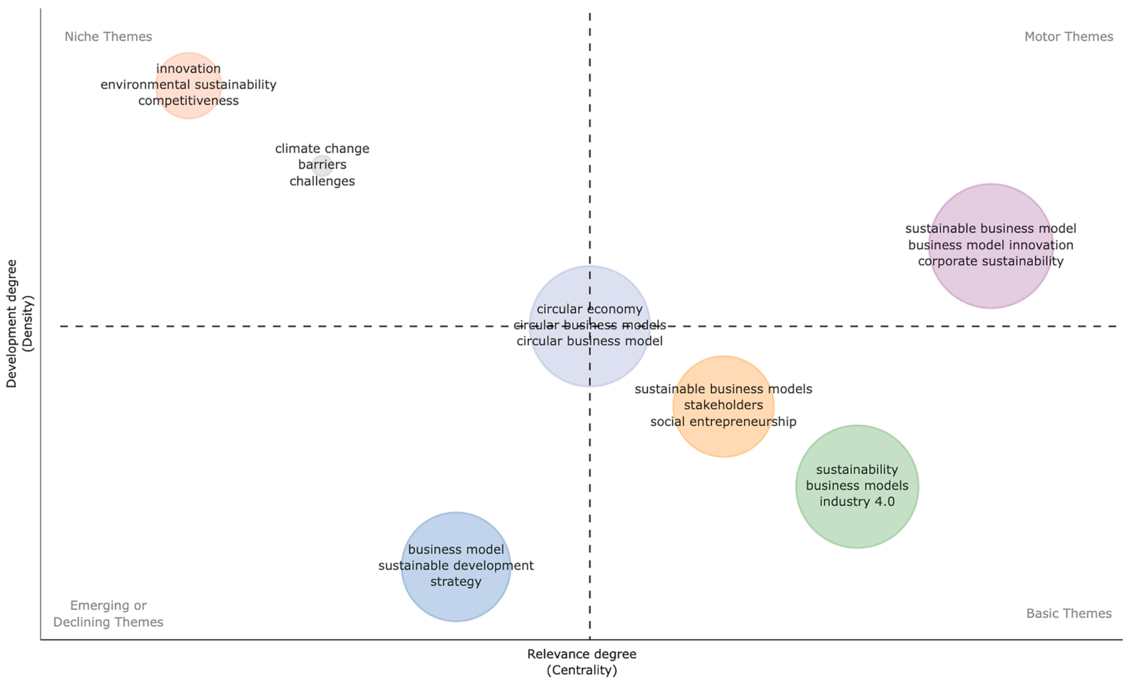


Fig. 9. Strategic thematic map – Second period (2014–2019).

To deepen this analysis, a qualitative review of key contributions within each thematic cluster is recommended, along with a focused keyword co-occurrence analysis to explore the conceptual underpinnings and evolution of each topic.

7. Discussion

This paper introduces Biblioshiny as a workflow-compliant implementation of the SAAS framework for transparent and reproducible science mapping. To empirically validate the effectiveness of this integrated approach, we replicated a published bibliometric study (Najmaei & Sadeghinejad, 2023) using Biblioshiny's unified environment. The discussion is organized in two complementary sections: first, we examine how the replication exercise validates key methodological claims about workflow integration; second, we position Biblioshiny as a practical solution to persistent challenges in bibliometric research.

7.1. Empirical validation through replication

The replication of Najmaei and Sadeghinejad's (2023) study serves as an empirical test of whether a workflow-integrated implementation can deliver on its promises of improved accuracy, efficiency, and reproducibility. The original study employed a fragmented multi-tool approach, with each analytical step requiring data export, format conversion, and re-import. Our implementation consolidated these operations within Biblioshiny's single-environment workflow, enabling direct assessment of the practical implications of workflow integration.

The most striking empirical evidence concerns data quality and preprocessing efficiency. While the original authors conducted three weeks of manual cleaning using BibExcel and CRExplorer, identifying approximately 330 citation variants, Biblioshiny's automated reference matching algorithm detected 9748 variants (24.4% of total references) with automation and minimal manual intervention. This increase is not merely quantitative but reveals systematic redundancy in bibliographic datasets that fragmented workflows often fail to capture. Critically, these normalized references are reusable across all subsequent analyses within the workflow - RPYS, co-citation analysis, and historiography - whereas tool-specific outputs from CRExplorer remained confined to single-use applications. This demonstrates that workflow integration does not simply aggregate functionalities but enables cumulative data quality improvements that propagate throughout the analytical pipeline.

The replication also validated core substantive findings while revealing how parameter transparency affects interpretation. Both studies identified an S-shaped diffusion pattern for the field, but our logistic growth model ($R^2 = 0.989$) projected maturity around 2027 rather than 2022. This difference is methodologically traceable: Biblioshiny's model uses deterministic fitting with documented parameters. The ability to inspect, document, and compare such choices illustrates how workflow compliance transforms methodological differences from opaque discrepancies into interpretable analytical decisions. Similarly, Reference Publication Year Spectroscopy confirmed identical foundational peaks (Brundtland, 1987; Carson, 1962; Teece, 2010) while additionally classifying references into trajectory categories (Hot Papers, Sleeping Beauties, Life Cycle, Constant Performers) through automated citation pattern analysis, an output unavailable in the original fragmented workflow.

The co-word analysis revealed both convergence and divergence. While the overall thematic trajectory (early sustainability → later circular economy) was reproduced, cluster positioning differed: where the original found an empty “motor theme” quadrant in 2014–2019, our analysis identified Circular Economy, Business Model Innovation, and Sustainable Business Models as motor themes. These differences are attributable to updated temporal windows, improved keyword disambiguation through Biblioshiny's functions, and transparent clustering algorithm selection (Louvain with documented modularity scores). Importantly, the ability to trace these divergences to specific parameter choices exemplifies how workflow transparency supports not just replication but methodological learning: readers can assess whether differences reflect substantive insights or analytical artifacts.

Across all modules, the replication confirms that an integrated workflow can achieve equivalent or superior analytical depth compared to multi-tool approaches while substantially improving efficiency, reproducibility (all parameters documented and exportable), and interpretability (consistent data normalization across analyses). This empirical validation supports the core claim that workflow compliance is not merely a convenience feature but a methodological necessity for credible bibliometric research.

7.2. Biblioshiny as a methodological solution

Recent reviews of bibliometric tools and practices have identified three persistent methodological gaps that limit the field's maturity as a cumulative, evidence-based discipline (Moral-Muñoz et al., 2020; Tomaszewski, 2023). First, workflow fragmentation forces researchers to combine partial functionalities across platforms - uploading to databases, exporting to visualization tools (VOSviewer, Gephi), returning to statistical environments (R, Python), and using separate tools for specialized tasks (CRExplorer for citation matching, LogletLab for growth modeling). Each transition introduces risks of data loss, format incompatibility, and undocumented parameter changes. Second, parameter opacity characterizes much published research: key methodological choices such as similarity thresholds, clustering algorithms, and normalization indices are frequently under-reported or described ambiguously, undermining reproducibility. Third, accessibility barriers create a stratified ecosystem where transparent, sophisticated analysis remains confined to programming-literate researchers, while non-technical audiences must rely on proprietary platforms with limited auditability.

Biblioshiny addresses these gaps through workflow compliance rather than technical superiority. The SAAS framework structures analysis as an explicit sequence where outputs from each phase become inputs for the next, eliminating export/import cycles and preserving workflow state. A critical, and underappreciated, implication of this architecture is data reusability: metadata cleaned during the Search phase (including deduplicated references, normalized author names, and harmonized journal titles) is automatically available to all downstream Analysis modules. This contrasts with tool-specific outputs (e.g., CRExplorer-disambiguated references used exclusively for RPYS), which cannot be ported to other analyses without manual re-export and reformatting. All parameter choices are user-visible, adjustable, and automatically documented, making reproducibility a default feature rather than requiring additional effort. The graphical interface provides point-and-click access to Bibliometrix's R functions while maintaining algorithmic transparency - users can export equivalent R code for any analysis, supporting hybrid workflows where graphical exploration transitions to programmatic customization.

This design responds to calls for democratization of bibliometric research without sacrificing rigor. Universities and funding agencies increasingly require evidence-informed decision-making but often lack programming expertise; Biblioshiny enables standardized assessments (emerging themes, research gaps, institutional strengths) without specialized personnel. Similarly, systematic reviews increasingly incorporate bibliometric mapping, but technical barriers have limited adoption (Zupic & Čater, 2015). By lowering entry barriers while preserving methodological control, exposing complexity through progressive disclosure rather than oversimplification, Biblioshiny supports wider diffusion of science mapping practices.

The SAAS implementation also addresses the standardization challenge identified by Cobo et al. (2011) and Moral-Muñoz et al. (2019): substantial heterogeneity in analytical choices across studies, combined with under-reporting, makes it difficult to distinguish substantive differences from methodological artifacts. Rather than imposing rigid prescriptions, SAAS enables "soft standardization" through structured reporting: by documenting parameters at each workflow stage, the framework supports methodological comparability while accommodating diverse research questions. When two studies employ different algorithms, both can report settings transparently, allowing readers to assess whether result differences reflect analytical choices or phenomena. This transforms methodological diversity from a reproducibility barrier into an opportunity for cumulative methodological learning.

8. Conclusion

This study introduces the SAAS framework as a principled methodological contribution to the field of science mapping, designed to make bibliometric workflows transparent, reproducible, and intellectually accountable. By articulating the research process as four explicit, sequential, and parameter-transparent stages, SAAS provides a conceptual scaffold that goes beyond the specificities of any individual software platform, defining what a rigorous science mapping workflow should entail independently of its operational implementation. In this sense, SAAS offers a structure through which principles of evidence synthesis, including those formalized in PRISMA, can be systematically applied within bibliometric analyses.

Biblioshiny serves as the empirical vehicle for demonstrating and validating the SAAS framework, offering a concrete instantiation of its principles within a unified and accessible environment. The replication exercises confirm that a SAAS-compliant implementation supports the systematic capture of methodological decisions and the reproducibility of science mapping outputs under equivalent data sources and parameter settings. More broadly, workflow compliance is positioned not as a feature of a particular tool, but as a methodological principle relevant to any software aiming to support credible bibliometric research.

Some limitations should nonetheless be acknowledged. First, as a memory-resident R application, Biblioshiny's performance depends on available RAM: very large collections (indicative of above 100,000 documents) may require high-memory workstations or corpus partitioning strategies. Second, the community detection step relies on the Louvain algorithm, whose outcomes are sensitive to the resolution parameter: different values can yield coarser or finer partitions of the keyword co-occurrence network, influencing the granularity of thematic clusters. Users are encouraged to assess the robustness of their clustering solutions by varying this parameter and by comparing results with those produced by alternative community detection algorithms implemented in Biblioshiny. Third, the validity of any bibliometric analysis is bounded by the quality of metadata retrieved from upstream sources, whether commercial databases such as Web of Science and Scopus or open infrastructures such as OpenAlex and Crossref, where missing fields, inconsistent disambiguation, and heterogeneous indexing practices can affect downstream outputs. The SAAS framework helps surface these dependencies by making the Search and Appraisal stages explicit and auditable, while Biblioshiny further supports users through reference matching, collection merging, and descriptive diagnostics. Critical judgment on data quality, however, remains an irreducible responsibility of the researcher.

Beyond this specific implementation, the SAAS framework opens several avenues for future development, including alignment with PRISMA principles, the design of interoperability protocols across platforms, and integration within emerging initiatives on responsible research assessment and open bibliometrics. As bibliometrics evolves toward greater transparency and democratized evaluation, workflow-compliant frameworks like SAAS offer a practical path forward, one where reproducibility is the default and rigorous analysis is accessible to the full spectrum of researchers, evaluators, and policymakers who depend on bibliometric evidence.

CRediT authorship contribution statement

Massimo Aria: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Corrado Cuccurullo:** Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Luca D'Aniello:** Writing – review & editing, Visualization, Formal analysis. **Maria Spano:** Writing – review & editing, Visualization, Formal analysis.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Appendix

The appendix A provides step-by-step instructions for installing the Bibliometrix R package and launching the Biblioshiny app; appendix B, instead, provides two complementary views of Biblioshiny's analytical capabilities: Table A1 presents the high-level SAAS workflow implementation across modules, while Table A2 documents the complete set of configurable parameters for reproducible science mapping analyses.

Appendix A

Installing and Launching Biblioshiny

The installation process requires a working R environment (version $\geq$ 4.5.0) but does not require prior programming experience.

Prerequisites

Before installing the latest version of Biblioshiny, ensure that:

- R version 4.5.0 or higher is installed on your system (download from https://www.r-project.org).
- (Optional) RStudio is installed for a more user-friendly interface (download from https://posit.co/products/open-source/rstudio/).
- An active internet connection is available for package installation.

Step 1: Install Bibliometrix package from CRAN.

Open R or RStudio and execute the following command in the console:

```
> install.packages("bibliometrix")
```

This command downloads and installs the Bibliometrix package and all its dependencies from the Comprehensive R Archive Network (CRAN). The installation process may take a few minutes depending on your internet connection speed.

Step 2: Launch Biblioshiny.

Once the installation is complete, load the Bibliometrix library and launch the Biblioshiny interface:

```
> library(bibliometrix)
> biblioshiny()
```

Upon execution, Biblioshiny will start a local web server and automatically open the application in your default web browser.

Step 3: Verify Successful Launch.

If the installation and launch were successful, the Biblioshiny home screen will appear in your browser, displaying the application interface as shown in Fig. A.1.

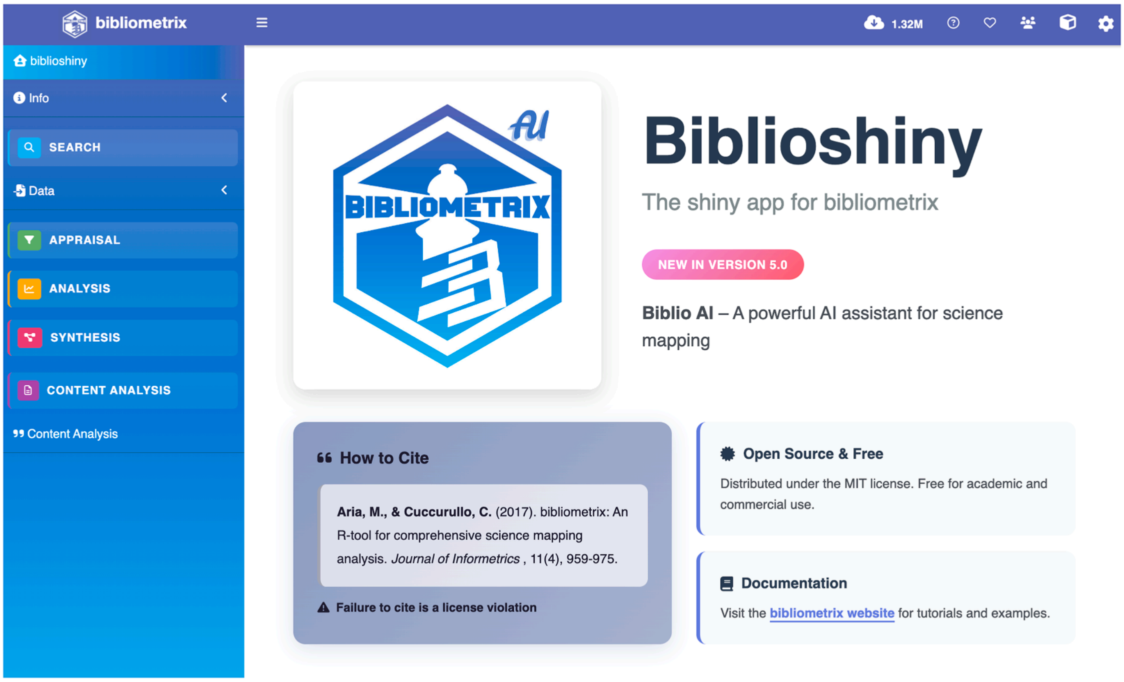


Fig. A.1. Biblioshiny home page.

Appendix B

Table B.1
SAAS Workflow Implementation in Biblioshiny.

SAAS Phase	Biblioshiny Module	Primary Functions	Key Capabilities
SEARCH	Data	Data import and integration	• Multi-source import (WoS, Scopus, Dimensions, PubMed, OpenAlex, Lens, Cochrane) • File format support (.bib, .csv, .txt, .RIS, .xlsx, .ciw) • API-based retrieval via OpenAlex and PubMed modules • Multi-database merging (Merge Collections) • Duplicate detection and removal • Automatic database recognition and field normalization
	Data – Pre-processing (Reference Matching)	Data cleaning and normalization	• Automated cited-reference matching and disambiguation • String-similarity algorithms (Jaro-Winkler, Levenshtein, OSA, LCS) • Author name standardization • Journal name normalization (ISO 4) • Affiliation disambiguation • Keyword normalization and synonym handling
APPRAISAL	Filters	Dataset refinement and screening	• Temporal filtering (publication year range) • Document type selection • Language filtering • Subject category (WC) filtering • Journal / source selection • Author and affiliation filtering • Geographical filtering (region and country) • Citation filtering (Total Citations and Citations per Year)
	PRISMA Diagram	Transparent screening documentation	• Interactive PRISMA 2020 flow diagram • Automatic record counts at each screening stage • Customizable inclusion/exclusion reasons • Export-ready diagram for systematic reviews
ANALYSIS	Overview	Descriptive analytics	• Main Information (collection summary statistics) • Annual Scientific Production • Average Citations per Year • Life Cycle Analysis (LCA) with logistic model fitting • Three-field plot (Sankey) • Collaboration indices
	Sources	Journal-level metrics	• Most Relevant Sources • Most Local Cited Sources • Bradford's law analysis • Source Local Impact (h, g, m-index, Total Citations) • Sources' Production over Time
	Authors	Author-, affiliation- and country-level metrics	• Author Profile • Most Relevant Authors (documents, % documents, fractionalized) • Most Local Cited Authors • Authors' Production over Time • Lotka's law • Authors' Local Impact (h, g, m-index, Total Citations) • Most Relevant Affiliations and Affiliations' Production over Time • Corresponding Author's Countries • Countries' Scientific Production and Production over Time • Most Cited Countries
	Documents	Document- and word-level metrics	• Most Global Cited Documents • Most Local Cited Documents • Most Local Cited References • References Spectroscopy (RPYS) • Citation trajectory classification (Thor et al., 2018) • Most Frequent Words • WordCloud • TreeMap • Words' Frequency over Time • Trend Topics
SYNTHESIS	Conceptual Structure	Co-word and thematic analysis	• Co-occurrence Network (keywords, titles, abstracts, subject categories)

(continued on next page)

Table B.1 (*continued*)

SAAS Phase	Biblioshiny Module	Primary Functions	Key Capabilities
			• Thematic Map (centrality × density, Callon et al., 1991) • Thematic Evolution (Sankey diagram across time slices) • Factorial Analysis (MCA, CA, MDS) with document clustering • Multiple clustering algorithms (Louvain, Leiden, Walktrap, Infomap, …) • Stopwords and synonyms customization
	Intellectual Structure	Co-citation analysis	• Reference Co-citation Network • Author Co-citation Network • Source Co-citation Network • Historiograph (direct citation network) • Multiple clustering and layout algorithms
	Social Structure	Collaboration network analysis	• Author Collaboration Network • Institution Collaboration Network • Country Collaboration Network • Countries' Collaboration World Map • Normalization (association, Jaccard, Salton, inclusion, equivalence) • Multiple clustering and layout algorithms
ADDITIONAL MODULES	Report	Results integration and export	• Interactive HTML/Excel report generation • Graphics and tables export (PNG, SVG, XLSX, CSV) • Automatic parameter documentation
	TALL Export	Standardized output for text analysis	• Export of bibliographic metadata for TALL (Text Analysis for All, www.tall-app.com) • Compatible format for downstream text-analysis workflows
	Content Analysis	Deep, AI-enhanced analysis of individual scientific articles	• Full-text PDF import and parsing • Citation extraction with surrounding context • Citation-network visualization at sentence level • Trend tracking of user-selected terms across article segments • AI-assisted summarization and interpretation (Biblio AI)
	Biblio AI	AI-assisted interpretation of results	• Natural-language explanation of analysis outputs • Context-aware commentary for tables and plots • Supports multiple LLM providers

Table B.2
Detailed Parameter Settings for Key Bibliometric Analyses.

Analysis Type	Technique	Parameter	Options / Range	Default
DATA CLEANING	**Reference matching**	Matching algorithm	Jaro-Winkler, Levenshtein, OSA, LCS	Jaro-Winkler
		Similarity threshold	0.70–0.98	0.85
		Keep original references	TRUE / FALSE	TRUE
		Add matching statistics	TRUE / FALSE	TRUE
		Export format	xlsx, rdata, both	xlsx
FILTERING	**Temporal filter**	Publication year range	Min–Max years of the collection	Full range
	Document type	Document type selection	Article, Review, Proceedings, Book, …	All
	Language	Language selection	Multi-selection	All
	Subject category	Subject category (WC)	Multi-selection	All
	Geographical filter	Region selection	Africa, Asia, Europe, N. America, S. America, Oceania, Seven Seas, Unknown	All
		Country selection	Multi-selection of countries	All
	Citation filter	Total Citations range	0–500	0–500
		Total Citations per Year range	0–100	0–100
OVERVIEW	**Life Cycle Analysis**	Growth model fitting	Logistic (automatic)	Automatic
		Saturation point (K)	Estimated automatically	Auto
	Three-field plot (Sankey)	Left field	AU, AU_UN, AU_CO, DE, ID, KW_Merged, TI, AB, SO, CR, CR_SO	CR
		Middle (central) field	AU, AU_UN, AU_CO, DE, ID, KW_Merged, TI, AB, SO, CR, CR_SO	AU
		Right field	AU, AU_UN, AU_CO, DE, ID, KW_Merged, TI, AB, SO, CR, CR_SO	KW_Merged (All Keywords)
		Top N items per field	1–50	20
SOURCES	**Most relevant sources**	Top N sources	1–100	10
	Bradford's law	Core / zones partitioning	Fixed (3 zones, Bradford law)	Fixed
	Source local impact	Impact metric	h-index, g-index, m-index, Total Citations	h-index
		Top N sources	1–100	10
AUTHORS	**Most relevant authors**	Frequency measure	N. of Documents, % of Documents, Fractionalized	N. of Documents

(*continued on next page*)

Table B.2 (*continued*)

Analysis Type	Technique	Parameter	Options / Range	Default
		Top N authors	1–100	10
	Most local cited authors	Top N authors	1–100	10
	Authors' production over time	Top N authors	1–100	10
	Lotka's law	Exponent (β) estimation	Automatic (least squares)	Auto
	Author local impact	Impact metric	h-index, g-index, m-index, Total Citations	h-index
		Top N authors	1–100	10
DOCUMENTS	**Reference Spectroscopy (RPYS)**	Reference separator	';', '.', ','	';'
		Year range (min–max)	Full CR year range	Full
		Median smoothing window	Centered, Backward	Backward
	Reference classification	Influential class (Thor 2018)	Constant Performer, Hot Paper, Life Cycle, Sleeping Beauty, Not Influent	Hot Paper
	Most cited documents	Citation type	Local (CR), Global (TC)	Global
CONCEPTUAL STRUCTURE	**Co-word network**	Field	ID (Keywords Plus), DE (Author's Keywords), KW_Merged, TI, AB, WC (Subject Categories)	KW_Merged
		N-grams (for TI/AB)	Unigrams (1), Bigrams (2), Trigrams (3)	Unigrams
		Stopwords / Synonyms files	User-loaded lists	None
		Min edges (co-occurrences)	≥ 0	2
		Top N nodes	5–1000	50
		Normalization	none, association, jaccard, salton, inclusion, equivalence	association
		Clustering algorithm	none, edge_betweenness, infomap, leading_eigen, leiden, louvain, spinglass, walktrap	louvain
		Remove isolated nodes	yes / no	yes
		Layout algorithm	auto, circle, fruchterman, kamada, mds, sphere, star	auto
		Node repulsion	0–1	0.5
		Node shape	box, circle, dot, ellipse, square, text	dot
		N. of labels	0–1000	1000
		Label size	0–20	3
		Edge size / transparency (α)	0–1	0.7
	Thematic Map	Field	ID, DE, KW_Merged, TI, AB	KW_Merged
		N-grams (for TI/AB)	Unigrams (1), Bigrams (2), Trigrams (3)	Unigrams
		Stemming	TRUE / FALSE	FALSE
		Stopwords / Synonyms files	User-loaded lists	None
		Min cluster frequency (per thousand docs)	1–100	5
		Number of words (Top N)	50–5000	250
		Clustering algorithm	none, edge_betweenness, infomap, leading_eigen, leiden, louvain, spinglass, walktrap	louvain
		α (centrality/density weighting)	0–1	0.5
		Node repulsion	0–1	0.5
		Label size	0–1	0.3
		Number of labels per cluster	0–10	3
	Thematic Evolution	Field	ID, DE, KW_Merged, TI, AB	KW_Merged
		N-grams (for TI/AB)	Unigrams (1), Bigrams (2), Trigrams (3)	Unigrams
		Stopwords / Synonyms files	User-loaded lists	None
		Number of cutting points	1–4	1
		Cutting point year(s)	User-defined year(s)	Mid-range
		Number of words (Top N)	50–5000	250
		Min cluster frequency (per thousand docs)	1–100	5
		α (occurrence vs. centrality)	0–1	0.5
		Min weight index (flow threshold)	0.02–1	0.1
		Clustering algorithm	none, edge_betweenness, infomap, leading_eigen, leiden, louvain, spinglass, walktrap	louvain
		Number of labels per cluster	1–5	3
		Label size	0–1	0.3
INTELLECTUAL STRUCTURE	**Co-citation network**	Unit of analysis	Cited references (CR), Authors (CR_AU), Sources (CR_SO)	CR
		Reference separator	';', '.', ','	';'
		Min edges (co-citations)	≥ 0	2
		Top N nodes	5–1000	50
		Clustering algorithm	none, edge_betweenness, infomap, leading_eigen, leiden, louvain, spinglass, walktrap	louvain
		Remove isolated nodes	yes / no	yes
		Layout algorithm	auto, circle, fruchterman, kamada, mds, sphere, star	auto
		Node repulsion	0–1	0.5

(*continued on next page*)

Table B.2 (*continued*)

Analysis Type	Technique	Parameter	Options / Range	Default
		Node shape	box, circle, dot, ellipse, square, text	dot
		N. of labels	0–1000	1000
		Label size	0–20	2
		Edge size	0.5–20	2
	Historiograph	Top N nodes	5–100	20
		Node label type	Short, Title, Keywords, Keywords Plus	Short
		Remove isolated nodes	yes / no	yes
		Node size	0–20	2
		Label size	0–20	3
SOCIAL STRUCTURE	**Collaboration network**	Unit of analysis	Authors (COL_AU), Institutions (COL_UN), Countries (COL_CO)	Authors
		Filter max authors per document	TRUE / FALSE	FALSE
		Min edges (collaborations)	≥ 0	1
		Top N nodes	5–1000	50
		Normalization	none, association, jaccard, salton, inclusion, equivalence	association
		Clustering algorithm	none, edge_betweenness, infomap, leading_eigen, leiden, louvain, spinglass, walktrap	louvain
		Remove isolated nodes	yes / no	yes
		Layout algorithm	auto, circle, fruchterman, kamada, mds, sphere, star	auto
		Node repulsion	0–1	0.5
		Node shape	box, circle, dot, ellipse, square, text	dot
		N. of labels	0–1000	1000
		Label size	0–20	2
		Edge size	0.5–20	5
		Edge transparency (α)	0–1	0.7

Note: All parameters listed in Table B2 are user-visible, adjustable, documented in downloadable reports, and replicable. This comprehensive parameter control distinguishes workflow-compliant implementations from tools where analytical choices remain implicit or inaccessible to users, thereby supporting the transparency and reproducibility objectives of the SAAS framework.

References

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open, 9*(1), 1–17. https://doi.org/10.1177/2158244019829575

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Aria, M., & Cuccurullo, C. (2026). *Science mapping analysis - A primer with biblioshiny*. McGraw-Hill. ISBN: 978-88-386-2297-7.

Bales, M. E., Wright, D. N., Oxley, P. R., & Wheeler, T. R. (2020). Bibliometric visualization and analysis software: State of the art, workflows, and best practices. https://hdl.handle.net/1813/69597.

Bankar, R. S., & Lihitkar, S. R. (2019). Science mapping and visualization tools used for bibliometric and scientometric studies: A comparative study. *Journal of Advancements in Library Sciences, 6*(1), 382–394.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Baxter, R., Christen, P., & Churches, T. (2003). A comparison of fast blocking methods for record linkage. In *Proceedings of the ACM SIGKDD Workshop on Data Cleaning, Record Linkage, and Object Consolidation* (pp. 25–27).

Beagrie, N., & Houghton, J. (2014). The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres. *Jisc*.

Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research, 67*(8), 1595–1600. https://doi.org/10.1016/j.jbusres.2013.10.001

Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P., & Fienberg, S. (2003). Adaptive name matching in information integration. *IEEE Intelligent Systems, 18*(5), 16–23. https://doi.org/10.1109/MIS.2003.1234765

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies, 1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), Article P10008.

Bocken, N. M. P., Short, S. W., Rana, P., & Evans, S. (2014). A literature and practice review to develop sustainable business model archetypes. *Journal of Cleaner Production, 65*, 42–56. https://doi.org/10.1016/j.jclepro.2013.11.039

Boeker, M., Vach, W., & Motschall, E. (2013). Google Scholar as replacement for systematic literature searches: Good relative recall and precision are not enough. *BMC Medical Research Methodology, 13*(1), Article 131. https://doi.org/10.1186/1471-2288-13-131

Boons, F., & Lüdeke-Freund, F. (2013). Business models for sustainable innovation: State-of-the-art and steps towards a research agenda. *Journal of Cleaner Production, 45*, 9–19. https://doi.org/10.1016/j.jclepro.2012.07.007

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*(1), 179–255. https://doi.org/10.1002/aris.1440370106

Börner, K., Sanyal, S., & Vespignani, A. (2007). Network science. *Annual Review of Information Science and Technology, 41*(1), 537–607. https://doi.org/10.1002/aris.2007.1440410119

Bornmann, L., & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics, 98*(1), 487–509. https://doi.org/10.1007/s11192-013-1161-y

Bota-Avram, C. (2023). Bibliometrics performance analysis. *Science mapping of digital transformation in business: A bibliometric analysis and research outlook* (pp. 15–22). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-26765-9_3

Brundtland, G. H. (1987). *Our common future/world commission on environment and development*. Oxford University Press.

Cagan, R. (2013). The San Francisco declaration on research assessment. *Disease Models & Mechanisms, 6*(4), 869–870. https://doi.org/10.1242/dmm.012955

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics, 22*(1), 155–205.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of Management Review, 4*(4), 497. https://doi.org/10.2307/257850

Carson, R. (1962). *Silent spring*. Houghton Mifflin.

Chen, C. (2016). *CiteSpace: A practical guide for mapping scientific literature* (pp. 41–44). Nova Science Publishers.

Chesbrough, H. (2010). Business model innovation: Opportunities and barriers. *Long Range Planning, 43*(2–3), 354–363. https://doi.org/10.1016/j.lrp.2009.07.010

Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change, 11*(3), 529–555. https://doi.org/10.1093/icc/11.3.529

Christen, P. (2012). *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-31164-2

Christen, P., & Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. *Quality measures in data mining* (pp. 127–151). Springer. https://doi.org/10.1007/978-3-540-44918-8_6

Ciampi, F., Demi, S., Magrini, A., Marzi, G., & Papa, A. (2021). Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation. *Journal of Business Research, 123*, 1–13. https://doi.org/10.1016/j.jbusres.2020.09.023

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics, 5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology, 63*(8), 1609–1630. https://doi.org/10.1002/asi.22688

Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. In *Proceedings of IJCAI-03 Workshop on Information Integration* (pp. 73–78). IIWeb.

Cuccurullo, C., Aria, M., & Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics, 108*(2), 595–611. https://doi.org/10.1007/s11192-016-1948-8

Diez-Junguitu, D., & Pena-Cerezo, M. A. (2026). BiblioMerge: A Python-based automated tool to merge WoS and Scopus bibliographic data, compatible with Biblioshiny, BibExcel, VOSviewer, SciMAT and ScientoPy. *Journal of Informetrics, 20*(1), Article 101770. https://doi.org/10.1016/j.joi.2026.101770

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review, 20*(1), 65. https://doi.org/10.2307/258887

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society, 20*(5), 325–344. https://doi.org/10.1080/01972240490507974

Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*(328), 1183–1210. https://doi.org/10.1080/01621459.1969.10501049

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139192675

Gan, Y. N., Li, D. D., Robinson, N., & Liu, J. P. (2022). Practical guidance on bibliometric analysis and mapping knowledge domains methodology-A summary. *European Journal of Integrative Medicine, 56*, Article 102203. https://doi.org/10.1016/j.eujim.2022.102203

Gärtner, A., Leising, D., & Schönbrodt, F. D. (2024). Towards responsible research assessment: How to reward research quality. *PLoS Biology, 22*(2), Article e3002553. https://doi.org/10.1371/journal.pbio.3002553

Gasparyan, A. Y. (2013). Bibliographic databases: Some critical points. *Journal of Korean Medical Science, 28*(6), Article 799. https://doi.org/10.3346/jkms.2013.28.6.799

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., & Hultink, E. J. (2017). The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production, 143*, 757–768. https://doi.org/10.1016/j.jclepro.2016.12.048

Gingras, Y. (2016). *Bibliometrics and research evaluation: Uses and abuses*. MIT Press.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of Management Review, 20*(4), 986. https://doi.org/10.2307/258963

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *PNAS, 112*(24), 7426–7431.

Klarin, A. (2024). How to conduct a bibliometric content analysis: Guidelines and contributions of content co-occurrence or co-word literature reviews. *International Journal of Consumer Studies, 48*(2), Article e13031. https://doi.org/10.1111/ijcs.13031

Kumar, R. (2025). Bibliometric analysis: Comprehensive insights into tools, techniques, applications, and solutions for research excellence. *Spectrum of Engineering and Management Sciences, 3*(1), 45–62. https://doi.org/10.31181/sems31202535k

Leydesdorff, L., Bornmann, L., Comins, J. A., & Milojević, S. (2016). Citations: Indicators of quality? The impact fallacy. *Frontiers in Research Metrics and Analytics, 1*, 1. https://doi.org/10.3389/frma.2016.00001

Magretta, J. (2002). Why business models matter. *Harvard Business Review, 80*(5), 86–92.

Matorevhu, A. (2024). Bibliometrics: Application opportunities and limitations. *Bibliometrics-An essential methodological tool for research projects. intechopen*. https://doi.org/10.5772/intechopen.1005292

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics, 106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Mont, O. K. (2002). Clarifying the concept of product-service system. *Journal of Cleaner Production, 10*(3), 237–245. https://doi.org/10.1016/S0959-6526(01)00039-7

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban Espejo, A. L., & Cobo Martín, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El Profesional de la Información, 29*(1), Article e290103. https://doi.org/10.3145/epi.2020.ene.03

Moral-Muñoz, J. A., López-Herrera, A. G., Herrera-Viedma, E., & Cobo, M. J. (2019). Science mapping analysis software tools: A review. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 159–185). Springer. https://doi.org/10.1007/978-3-030-02511-3_7.

Najmaei, A., & Sadeghinejad, Z. (2023). Green and sustainable business models: Historical roots, growth trajectory, conceptual architecture and an agenda for future research - A bibliometric review of green and sustainable business models. *Scientometrics, 128*(2), 957–999. https://doi.org/10.1007/s11192-022-04577-2

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), Article 026113.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Wiley.

Paskin, N. (2005). Digital object identifiers for scientific data. *Data Science Journal, 4*, 12–20. https://doi.org/10.2481/dsj.4.12

Pasula, H., Marthi, B., Milch, B., Russell, S. J., & Shpitser, I. (2002). Identity uncertainty and citation matching. *Advances in Neural Information Processing Systems*, 15.

Põder, E. (2022). What is wrong with the current evaluative bibliometrics? *Frontiers in Research Metrics and Analytics, 6*, Article 824518. https://doi.org/10.3389/frma.2021.824518

Porter, M. E. (1979). How competitive forces shape strategy. *Readings in strategic management* (pp. 133–143). Macmillan Education UK.

Porter, M. E., & van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *Journal of Economic Perspectives, 9*(4), 97–118. https://doi.org/10.1257/jep.9.4.97

Richardson, J. (2008). The business model: An integrative framework for strategy execution. *Strategic Change, 17*(5–6), 133–144. https://doi.org/10.1002/jsc.821

Schaltegger, S., Freund, F. L., & Hansen, E. G. (2012). Business cases for sustainability: The role of business model innovation for corporate sustainability. *International Journal of Innovation and Sustainable Development, 6*(2), 95. https://doi.org/10.1504/ijisd.2012.046944

Shahrzadi, L., Mansouri, A., Alavi, M., & Shabani, A. (2024). Causes, consequences, and strategies to deal with information overload: A scoping review. *International Journal of Information Management Data Insights, 4*(2), Article 100261. https://doi.org/10.1016/j.jjimei.2024.100261

Singh, D., & Garg, R. (2025). A survey of community detection algorithms and its comparative performance analysis. *Computer Science Review, 58*, Article 100799. https://doi.org/10.1016/j.cosrev.2025.100799

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science, 50*(9), 799–813. https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID−SI9>3.0.CO;2-G

Stubbs, W., & Cocklin, C. (2008). Conceptualizing a "sustainability business model. *Organization and Environment, 21*(2), 103–127. https://doi.org/10.1177/1086026608318042

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43*(2–3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003

Thor, A., Bornmann, L., Marx, W., & Mutz, R. (2018). Identifying single influential publications in a research field: New analysis opportunities of the CRExplorer. *Scientometrics, 116*(1), 591–608. https://doi.org/10.1007/s11192-018-2733-7

Tomaszewski, R. (2023). Visibility, impact, and applications of bibliometric software tools through citation analysis. *Scientometrics, 128*(7), 4007–4028. https://doi.org/10.1007/s11192-023-04725-2

Tukker, A. (2015). Product services for a resource-efficient and circular economy-A review. *Journal of Cleaner Production, 97*, 76–91. https://doi.org/10.1016/j.jclepro.2013.11.049

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36*(3), 397–420. https://doi.org/10.1007/BF02129602

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics, 59*(3), 467–472.

Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Muñoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management, 60*(4), Article 103385. https://doi.org/10.1016/j.ipm.2023.103385

Wagner, B., & Svensson, G. (2010). Sustainable supply chain practices: Research propositions for the future. *International Journal of Logistics Economics and Globalisation, 2*(2), 176. https://doi.org/10.1504/IJLEG.2010.032432

Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods* (pp. 354–359). American Statistical Association.

Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology, 66*(2), 408–427. https://doi.org/10.1002/asi.23179

Zott, C., Amit, R., & Massa, L. (2011). The business model: Recent developments and future research. *Journal of Management, 37*(4), 1019–1042. https://doi.org/10.1177/0149206311406265

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629